

```

1 MNHDIITFLTL FLLGFFGGTH CIGMCGGLSS AFALQPPHI NRFWLILLN
51 TGRISSTYAT GLMLGIGOL GISLDOTRVL QNLYTASNL LLFLGLYTS
101 GISSLAAKIE KIGKPIWRNL NPILNRLPI KSIPACLAVG ILWGLPCGL
151 VYSASLYALG SSGSATGGLY MIAFALGTLP NLLAIGIFSL QLKIMQNRV
5 IRLCTGLSVS LMAIWLAVL WL*
201

```

In addition, ORF103ng and ORF103-1 show 97.3% identity in 222 aa overlap:

```

10 orf103-1.pep      10      20      30      40      50      60
      MNHDIITFLTLFLLGFFGGTHCIGMCGGLSSAFALQPPHINRFWLILLNTGRVSSYTAI
      |||
10 orf103ng         10      20      30      40      50      60
      MNHDIITFLTLFLLGFFGGTHCIGMCGGLSSAFALQPPHINRFWLILLNTGRISSTYAI
      |||

15 orf103-1.pep      70      80      90      100     110     120
      GLILGLIGQVGSLSDDQTRVLQNILYTAANLLLFLGLYLSGSISSLAAKIEKIGKPIWRNL
      |||
15 orf103ng         70      80      90      100     110     120
      GLMLGIGQLGISLDDQTRVLQNILYTAANLLLFLGLYLSGSISSLAAKIEKIGKPIWRNL
      |||

20 orf103-1.pep      130     140     150     160     170     180
      NPILNRLPIKSIPACLAVGILWGLPCGLVYSASLYALGSGSATGGLYMLAFALGTLP
      |||
20 orf103ng         130     140     150     160     170     180
      NPILNRLPIKSIPACLAVGILWGLPCGLVYSASLYALGSGSATGGLYMLAFALGTLP
      |||

25 orf103-1.pep      190     200     210     220
      NLLAIGIFSLQLKIMQNRVIRLCTGLSVSLWALWKLAVLWLX
      |||
25 orf103ng         190     200     210     220
      NLLAIGIFSLQLKIMQNRVIRLCTGLSVSLWALWKLAVLWLX
      |||

```

30 Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 47

35 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 399>:

```

1 ATGGAAAACC AAAGGCCGCT CTAGGCTTT CGCTGGCAC TTTTGGCGGC
51 GATGACGTGG GSAACGCTGC CGAT.TCCGT CGCGCAGGTA TTGAAGTTTG
101 TCGATCGCGC GACGCTGGTG TGGCTGGCT TTACGCTGGC GCGCGCGGTA
151 TTGTTTGGTT TCGCTGGCACT GCGCGGAGC CTGCGGAGC GCGCGGATTG
40 201 TTTCTGGTG CTCAATTCAGG CTGCTGCTGC TCGCGGTGGC GGCATTCGTT
251 GCAACACTTG TCGTGATTGC CAAAGGCGCT CMTTATATT CGCGGACCAC
301 GACGCGAGTT TTGTGGCAGA TTTCGCGGTT TACGATGATT GTGTGCGGTG
351 TGTGTGGTGT TAAAGACCGG ATGACTGCGC CTCAAGAAAT CGGCTTGTTT
401 TTGCTGCTTG CCGGTTTGCT TATGTATTTT AACGATAAAT TCGGCGAGTT
45 451 TCGCGGTTTG GCGCGGTATG C.AAGGCGCT GTTGTGCTGT GCGCGAGGCA
501 GTATGGCAATG GGTGTGTAAT GCGCTGGCGC AAAAGCTGCT GTCGCGGCAA
551 TTGCGGCGGC AACAGATTCT GCTGTTGATT TATGCGGCAA GTGCGCGCTG
601 GTTCTGCGCG TTGCGCGAAC CGGCACACAT CGGAAGTATG GACGGTACGT
651 TGGCGTGGGT ATGTATTGGG TATTGCTGCT TGAATACGTT AATCGGTTAC
50 701 GGCTCGTTTG CGGAGGCGTT GAAACATTGG GAGGCTTACA AAGTCAGCGC
751 GGTAACAAAC TTGCTCCCGG TTTTACCGT AATAAATACT TTGCTCGGGC
801 ATTATGTGAT GCCTGAAACT TTGCGCGCGC CGGA..

```

This corresponds to the amino acid sequence <SEQ ID 400; ORF104>:

```

55 1 MENQRPLLGF RLALLAMTW GTPXSVRQV LKFVDAPTLV WVRFTVAADV
51 LFVLLALGSR LNRDRDFWC SFRLLLGVA GIGANFLVIA QGLHYISPTT
101 TQVLWGISIF THIVVGVLVF KRMRTAQKI GLVLLAGLI MYPNKFGEL
151 SGLGAYKKGV LILNAGSMW VCNAAVAKLL SAQFGPQOIL LLIYAASAV
201 FLFPAEPAHI GMDGTLAWV CIAYCCINTL IGYGSGFEAL KHWEASKVA

```


10

25

35

45

50

55

60

65

	orf104-1	 KDRMTAAQKIGLVLLLAGLIMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL 130 140 150 160 170 180
5	orf104a.pep	190 200 210 220 230 240 SAQFGPQOILLIYAASAAVFLPFAELAHIGSLDGTALAWVCFAYCCINTLIGYGSFGEAL
10	orf104-1	190 200 210 220 230 240 SAQFGPQOILLIYAASAAVFLPFAEFAHIGSLDGTALAWVCFAYCCINTLIGYGSFGEAL
	orf104a.pep	250 260 270 280 290 300 KHWEASKVSAVTTLLEPVFTVIFSLGHYVMPDTFAAPDMNGLGYAGALVVVGAVTAAG
15	orf104-1	250 260 270 KHWEASKVSAVTTLLEPVFTVIXXILGHYVMPETFAAP

Homology with a predicted ORF from *N.gonorrhoeae*

ORF104 shows 93.9% identity over a 277aa overlap with a predicted ORF (ORF104.ng) from *N. gonorrhoeae*:

20	orf104.pep	MENQRPLLGFRLLAALLAAMTGTLPKXSVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR	60
	orf104ng	MENQRPLLGFRLLAALLAAMTGTLPKXSVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR	60
25	orf104.pep	LPKRRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPTTMIWVGVLVF	120
	orf104ng	LPKRRDFSWHSFRLLLLGVGTISANFVLIAQGLHYISPTTTQVLWQISPTTMIWVGVLVF	120
30	orf104.pep	KDRMTAAQKIGLVLLLAGLIMYFNDKFGELSGLGAYAKGVLLCAAGSMAWVCNAVAQKLL	180
	orf104ng	KDRMTAAQKIGLVLLLAGLIMYFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL	180
	orf104a.pep	SAQFGPQOILLIYAASAAVFLPFAEFAHIGSLDGTALAWVCIAYCCINTLIGYGSFGEAL	240
	orf104ng	SAQFGPQOILLIYAASAAVFLPFAEFAHIGSLDGTALAWVCFVYCCINTLIGYGSFGEAL	240
35	orf104.pep	KHWEASKVSAVTTLLEPVFTVINTLLGHYVMPETFAAP	277
	orf104ng	KHWEASKVSAVTTLLEPVFTVIFSLGHYVMPDTFAAPDMNGLGYAGALVVVGAVTAAG	300

The complete length ORF104ng nucleotide sequence <SEQ ID 405> is predicted to encode a protein having amino acid sequence <SEQ ID 406>:

1	MENQRPLLGFRLLAALLAAMTGTLPKXSVRQVLKFVDAPTLVWVRFTVAAAV
51	LFVLLALGGR LPKRRDFSWH SFRLLLLGVTF GISANFVLIA QGLHYISPTT
101	TQVLWQISPTT TMIVVGVLFV KDRMTAAQKI GLVLLLVGLL MFFNDKFGEL
151	SLGAYAKGV LLCAAGSMAW VCYAVAQKLL SAQFGPQOIL LLIYAASAAV
201	FLKAEPAHI GSLDGTALAW CFVYCCINTL IGYGSFGEAL KHWEASKVSA
251	VTTLLPEFTV IFSLLGHYVM PDTFAAPDMN GLGYAGALVV VGVAVTAAG
301	DRPFKRR*

Further work revealed the complete gonococcal nucleotide sequence <SEQ ID 407>:

1	ATGGAARACC AAAGCGCGCT CCTGAGGCTC GCGTGGCGAC TTTTGGCGGC
51	GATGACGTCG GCACCGCTCG CCAATGCGGT CGACAGGTTA TTCAAGTTTG
101	TGATGCGCGC GACCGCTGCT TGGGTGCGTT TTACCGTGGC CGCGCGGTG
151	TGTGTTGTTT TGTCTGCATT GGGCGGCGCG CTGCGGAAGC GCGCGGATT
201	TTCTTGCCAT TCATTGAGC TGCTGCTGCT CGCGCTGAGC GGCATTTCGG
251	CAACATTTGT GCTGATTGCC CAAGGCGTGC ATTATATTTC CGCAGCACG
301	ACGCGAGTTT TGTGCGAGAT TTGCGCGTTC ACGATGATTG TTGTGCGCGT
351	GTTGCTGTTT AAAGACCGGA tgaCTGCGCG GCAGAAAATC GGTTTGTTT
401	TGCTGCTgtT CGGTTgtCTT ATGTTTtTa ACAGCAAAAT CGCGAGTTGT
451	TGCGGTTTGG GCGCGTATCG GAAGGCGGTG TTGCTGTGTG CGCGAGCGAG
501	TATGGCGTGG GTGCTGTTATG CGGTGCGGCA AAGCTGCTGT TCGCGCGCAAT
551	TGCGGCGGCA ACAGATTCTG CTGTTGATTT ATGCGGcaag tgccgcgcGTG
601	TTCTGccgT TTGCGgaACC GGCACACATC GGAAGTTTgg aCGGTACGTT
651	GCGCTGGGTG TGTTTTGTGT ATTGCTGCTT GAATACGTTA ATCGGTGATG

701 GCTCGTTCGG CGAGGCGGTTG AACATTGGG AGGCTTCCAA AGTCAGCGCG
 751 GTACAACTT TGCTGCCCGT GTTTACCGTA ATATTTTCTT TGCTCGGGCA
 801 TTATGTGATG CTTGATACATT TTGCCGCGCC GGATATGAAC GGTTCGGGT
 851 ATGTCCGCGC ACTGGTCGTG GTCGGGGGTG CGGTTACGCG GCGCGTGGGG
 901 GACAGGCCGT TCAAACGCGC CTAG

This corresponds to the amino acid sequence <SEQ ID 408; ORF104ng-1>:

1 MENORPLLGF ALALLAAMTW GTLPVAVRQV LKFVDAPTLV WVRFTVAARV
 51 LFVLLALGGR LPKRRDFSWH SFRLLLVGT GISANFVLA QGLHYISPTT
 101 TQVLWQISPF TMIVVGVLFV KDRMTAAQKI GLVLLVGLL MFFNDKFGEL
 151 SGLGAYAKGV LLCAAGSMW VCYAVAQKLL SAQFGPQOIL LLIIYAASAV
 201 FLPAEPAHI GSLDGTIAWV CFVYCCCLNT IGYGSGEAL KHWEASKVSA
 251 VTLLPVFTV IFSLLGHVYM PDTFAAPDMN GLGVGVALLV VGGAVTAARV
 301 DRPFKRR*

ORF104ng-1 and ORF104-1 show 97.5% identity in 277 aa overlap:

15	orf104-1.pep	10	20	30	40	50	60
	MENORPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAARVLFVLLALGGR						
	orf104ng-1	10	20	30	40	50	60
	MENORPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAARVLFVLLALGGR						
20	orf104-1.pep	70	80	90	100	110	120
	LPKRRDFSWCSFRLLLVGAGISANFVLAQGLHYISPTTQVLWQISPTTMIIVGVLFV						
	orf104ng-1	70	80	90	100	110	120
	LPKRRDFSWHCSFRLLLVGTGISANFVLAQGLHYISPTTQVLWQISPTTMIIVGVLFV						
25	orf104-1.pep	130	140	150	160	170	180
	KDRMTAAQKIGLVLLVGLMFFNDKFGELSGLGAYAKGVLLCAAGSMWVCYAVAQKLL						
	orf104ng-1	130	140	150	160	170	180
	KDRMTAAQKIGLVLLVGLMFFNDKFGELSGLGAYAKGVLLCAAGSMWVCYAVAQKLL						
30	orf104-1.pep	190	200	210	220	230	240
	SAQFGPQOILLIIYAASAAVFLPAEPAHIGSLDGTIAWVCFVYCCCLNTLIGYSGEAL						
	orf104ng-1	190	200	210	220	230	240
	SAQFGPQOILLIIYAASAAVFLPAEPAHIGSLDGTIAWVCFVYCCCLNTLIGYSGEAL						
35	orf104-1.pep	250	260	270	280	290	300
	KHWEASKVSAVTTLLPVFTVIXXLLGHVMPETFAAP						
	orf104ng-1	250	260	270	280	290	300
	KHWEASKVSAVTTLLPVFTVIFSLLGHVMPDFTAAPDMNGLGVGVALLVGGAVTAARV						

In addition, ORF104ng-1 shows significant homology with a hypothetical *H. influenzae* protein:

45 gi|1573895 (U32769) hypothetical [Haemophilus influenzae] Length = 306
 Score = 237 bits (598), Expect = 8e-62
 Identities = 114/280 (40%), Positives = 168/280 (59%), Gaps = 8/280 (2%)

50 Query: 30 QRPXXXXXXXXXXMTWGTLPVAVRQVLKFVDAPTLVWXXXXXXXXXXXXXXXXXXXXP 88
 Q=P M WG+LPVIA++QVL ++A T+VW P
 Sbjct: 3 QQPLLGTFALTITAMAGSLPIALKQVLSSNAQITIVYRFTIAAVSLALLAYKKQLPE 62

55 Query: 89 --KRRDFSWHCSFRLLLVGTGISANFVLAQGLHYISPTTQVLWQISPTTMIIVGVLFV 146
 R R ++W ++L+G V ++NE+L + L+YI P+ Q+ +S F M++ GVL+P
 Sbjct: 63 LMKVRQYAW----IMLIGVIGLTSNLFSSSLNIYEPSVAQTFIHLSSFGMLICGVLI 118

60 Query: 147 KDRMTAAQKIXXXXXXXXXXXMFFNDKFGELSGLGAYAKGVLLCAAGSMWVCYAVAQKLL 206
 K+++ QKI +FFND+ +GL Y+ GV+L G++ WV Y+AQKL+
 Sbjct: 119 KEKLLGKIGLIFLLILGLGLFFNDRDFAFAGLQYSTGVILGVGGALLWVAGMQLIM 178

65 Query: 207 SAQFGPQOILLIIYAASAAVFLPAEPAHIGSLDGTIAWVCFVYCCCLNTLIGYSGEAL 266
 +F QOILL++Y A F+P A+ + + L LA +CF+YCCCLNTLIGYGS+ EAL
 Sbjct: 179 LRFNFSQILIMYLGCAIAFMPMADFQVQELT-PLALICFTYCCCLNTLIGYSGYAEAL 237

Query: 267 KHWEASKVSAVTTLLPVFTVIFSLLGHVMPDFTAAPDMN 306

W+ SKVS V TL+P+ET++FS + HY P FAAP++N
 Sbjct: 238 NRWDVSKVSVVITLVPLFTILFSHIARYFSPADFAAPELN 277

Based on this analysis, including the presence of a putative leader sequence and several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 48

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 409>:

```

1 ATGGTAGCTC GTCGGGCTCA TAACCCGAAG GTCGTAGGTT CGAATCCTGT
51 .CCCGCAACC TAATTTCAAA CCCCTCGGTT CAATGCCGAG GG.GTTTTGT
101 T.TFGCCTGT TTCCTGTTTC CTGTTTCTGT CGCCTCCGTT TTTTTCGCGG
151 ATTTTCTCCTC CGGCCGCAAT ATCGGAACGG CAGACGCCCG CTCGTTTGGCG
201 GTTGCAAAAT CAGGCAAGTTT GGCTACAATC TTCGCATTG TCTTCAGAA
15 251 AGCCAAACCAT GCGCCACCGT CGTTPACCG AATCCGCTAG CAACAAGAC
301 CTTGATGCTC TGTTCGAGTG GGCMAAAGCA AGTTACGGTG CAGAAAGTTG
351 CTGGAACACG CTGATCTCTGA ACGGTCysCC TTTGGGCAAC CTGTCGCCGG
401 AATGGGTGGA ACGCGTsmmA AAAGACTGGG AGGCAGCGTG CyCGAGTCT
451 TCAGACGGCA TTTTCTGAA TgCGGACGGG TgGctGATA TGGGCGGAg
501 cTTACAGCAC CTCGCCCTCG GTTGGCACTG TGCGGGGCTG TTGAGCGsT
20 551 GGCGCAACGA GTGTTTTCGAC CTGACGACGG CGCGCGGCAA CCCCTTGTTC
601 ACGCTCGsac CGCGCGyTTT mCGTCTCTKTC GCACTGCTCA CGCGCGCGCT
651 CCACTCTCAC GGTCTGACCG AATCGGAGGG CGATGCGCAT TTCTGATAG
701 GCAAGGCGAG TCCGACAAJ CGAGTCSATC CCAACAACTC GCACATATCT
751 rCGCGCGGCG GTGTTTCCGG CGGGGAATG CCCTCTGAAG CCGTGTCTCG
25 801 CGAAGACAGC GAAGAAGCGG GTTTGATATA AACGCTcTT CCGCTCATCC
851 GCCCGGTATC GCAGCTGCAC AGCCTGCGCT CGCTCAGCGG GGGTGTACAC
901 AATGAATACC TGTATGTATT CGATCGCGCT CTGCGG...
```

This corresponds to the amino acid sequence <SEQ ID 410; ORF105>:

```

1 MVARRAHNPV VVGSNPXPAT XFQTPRFNAE XVLXLPVSCF LFPAASVFECR
30 51 IFLFAAISER QTAVCLRLLQI QAVWLQSSAL SSRKPTMPV RFTESVSKQD
101 LDALFEWAKA SYGAESCWKT LYINGXPLGN LSPFWVERVX KDWEAGCXES
151 SDGILFNADG WPDMGGRLLQH LALGWHCAGL LDGWRNECFD LTDGGNPLF
201 TLERAXXRPX GLLSRAVHLN GLTESDGRWH FWIGRRSPHK AVDPNKLNTF
251 XAGVGSGGEM PSEAVCRESS EEAGLDKTLT PLIRPVSQLH SLRSVSRGVH
35 301 NEILYFDDAV LP...
```

Further work revealed the complete nucleotide sequence <SEQ ID 411>:

```

1 ATGCGGACCG TCCGTTTTAC CGAATCCGCT AGCAAAACAG ACCTTGATGC
51 TCTGTTTCGAG TGGCAAAAG CAAGTTACGG TGCAGAAAGT TGCTGGAAAA
101 CGCTGTATCT GAACGGTCTG CCTTTTGGCA ACTGTTGCGC GGAATGGGGT
40 151 GAACGCGCTGA AAAAAGACTG GAGGCGAGCG TGCTCGAGAT CTTACAGACG
201 CATTTTCTCTG AATTCGGAGC CTGCGCGGCG TATGGGCGGA CGCTACACGC
251 ACCTCGCGCT CGGTGACGAC TGTCGGGCGG TGTTGGACGG CTGCGGCAAC
301 GAGTGTTCG ACCTGACCGA CGGCGGCGCG AACCCCTTGT TCAGCCTCGA
45 351 ACGCGCGGCT TTCGTCCTT TCGGACTGCT CAGCGCGGCC GTCCATCTCA
401 ACGTCTCGAC CGAATCGGAC GGCGGATGGC ATTTCTGAT AGGCAGGCGC
451 AGTCCGACACA AAGCAGTCTG TCCCAACAAA CTGCACAAAT CTGCCGCGCG
501 CGTGTGTTTCC GCGCGCGAAA TGCGTCTGA AGCGGTGTGT CGGGAAGACA
551 GCGAAGAAGC CGGTTTGGAT AAAACGCTGC TTCGCTCAT CGCCCGGGA
601 TCGCAGCTGC ACAGCTGCGG CTCGCTCAGC CGGGGCTGAC ACAATGAATC
50 651 CCTGTATGTA TTCGATGCGG TCTCTGCGGA AACCTTCTG CTTGAAATTC
701 AGGATGGCGA AGTGGCGGGT TTTGAGAAAA TGGACATCGG CGGTCTGTTG
751 GATGCCATGT TGTGCGGAAC CATGATGCAC GACGCGCAAC TGGTTACGCT
801 GGA CGCGGTTT TCGGTTACG GTCTGATTGA TGCGGCCCAT CCGCTGTCCG
851 AGTGGCTGGA CGGCATACGT TTATAG
```

This corresponds to the amino acid sequence <SEQ ID 412; ORF105-1>:

```

      1  MPTVRFTESV SKQDLDALE WAKASYGAES CWKTLYLNLG PLGNLSPEWV
    51  ERVKKDWEAG CSESSDGIFL NADGWPMGG RLQHLALGWH CAGLLDGRWN
   101  ECFDLTDCG NPLFLTERA FAEPLLSRA VRLNGLTESD GRWFHWIGRR
   151  SPHKAVDPNK LDMTAAGVS GGEMPEAVC RESSEAGLD KTLPLIRFV
    5  201  SQLHSLSRSVS RGVHNEILYV FDAVLPETFL PENQDGEVAG FEKMDIGLL
      251  DMLSGNMMH DAQLVTLDAF CRYGLIDAAH PLEWLDGIR L*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF105 shows 89.4% identity over a 226aa overlap with an ORF (ORF105a) from strain A of *N.*

meningitidis:

```

      60      70      80      90     100     110
orf105.pep  ISERQTAVCLLQIQAVNLQSSALSSRKPTMPTVRFTESVSKQDLDALEWAKASYGAES
      15  orf105a  |||||
               MPTVRFTESVSKHDLDALEWAKASYGAES
               10      20      30

      120     130     140     150     160     170
orf105.pep  CWKTLYLNGXPLGNLSPEWVRVKKDWEAGCESSDGIFLNADGWPMGGRRLQHLALGWH
      20  orf105a  |||||
               CWKTLYLNLGPLGNLSPEWAEVRVKKDWEAGCESSDGIFLNADGWPMGGRRLQHLARIWK
               40      50      60      70      80      90

      180     190     200     210     220     230
orf105.pep  CAGLLDGRWNECDFLTDGGGNPLFTLERAXRFXGLLSRAVHNLGLTESDGRWFHWIGRR
      25  orf105a  |||||
               EAGLLHGWDECDFLTDGGSNPLFALERAAPFPGLLSRAVHNLGLTESDGRWFHWIGRR
               100     110     120     130     140     150

      240     250     260     270     280     290
orf105.pep  SPHKAVDPNKLDNTXAGVSGGEMPEAVCRESSEAGLDKTLPLIRFVSQLSHSLSRSVS
      30  orf105a  |||||
               SPHKAVDPDKLDNTAAGVSSGELPSETVCRESSEAGLDKTLPLIRFVSQLSHSLSRSVS
               160     170     180     190     200     210

      300     310
orf105.pep  RGVHNEILYVFDVLP
      35  orf105a  |||||
               RGVHNEILYVFDVLPETFLPENQDGEVAGFEKMDIGLLAAMLSGNMMHDAQLVTLDAF
               220     230     240     250     260     270

```

40 The complete length ORF105a nucleotide sequence <SEQ ID 413> is:

```

      1  ATGCCGACCG TCCGTTTAC CGAATCCGTC AGCAACAACG ACCTTGATGC
    51  CCTATTTCGAG TGGGCAAAAG CAAGTTACGG TGGCGAAAGT TCGTGGAAAA
   101  CGCTGTATCTC GAACGGTCTG CTTTGGGGCA ATCTGTGCGC GGAATGGGGC
   151  GAGCGGCTCA AAAAAGACTG GGAGGCAGCG TGCTGGAGT CTTCAAGCGG
   201  CATTTTCCTG AATGGCGGAG GCTGGCCAGA TATGGCAGA CGCTTGACGC
   251  ACCTCGCCGC AATATGGAAA GAAGCGGGAC TGCTTACGCG CTGGCGGCGC
   301  GAGTGTTTTCG ACCTGACCCA CGCGCGCAGC AATCCCTGTG TCGGCGTCTGA
   351  ACGCGCCGCT TCCGCTGCGT TCGACATGCT GTCGCGCGCC GTCCATCTCA
   401  ACGGTTTGGT CGAATCGGAC GCGGATGCGC ATTTCTGGAT AGGCACGCGC
   451  AGTCTCCACA AAGCAGTCTG TCCCGACAAA CTCGACATA CTGCGCCGCG
   501  CGGTGTTTCC AGCGGTGAAT TCGCCTCTGA AACCGTGTGT CGCGAAGACA
   551  GCGAAGAAGC CGGTTTGGAT AAAACGCTGC TTCCGCTCAT CCGCCCGGTA
   601  TCGCAGCTCGC ACAGCCTGCG CCGCGTCAGC CGGGTGTGCG ACAATGAAAT
   651  CCTGTATGTA TCGATGCGG TCCTGCCCCA AACCTTCTGT CCTGAAATCT
   701  AGGATGCGGA AGTGGCGGGT TTTGAGAAAA TGGACATCGG CGGTCTGTGT
   751  GCTGCCATGT GTGCGGGAAA CATGATGCAC GACGCGCAAC TGGTTACGCT
   801  GGACGCGTTT TGCGGTACG GTCTGATTGA TGGCGCCCAT CCGCTGTGCG
   851  AGTGGCTGGA CGGCATACGT TTATAG

```

This encodes a protein having amino acid sequence <SEQ ID 414>:

```

    60  1  MPTVRFTESV SKHDLDALE WAKASYGAES CWKTLYLNLG PLGNLSPEWA
      51  ERVKKDWEAG CSESSDGIFL NADGWPMGGR RLQHLARIWK EAGLLHGRWD

```

```

101 ECFDLTDGGS NPLFALEAA FRPFGLLSRA VHLNGLVESD GRWHFWIGRR
151 SFHRAVDYK LDNTAGGVV SGLEPSETVC RESSEAGLID KTLLELTVV
201 SOLHSLRPVS RGVHNEILYV FDAVLPETFL PENQDGEVAG FEKMDIGGLL
251 AAMLSGNMH DAQVLTDAF CRYGLIDAAH PLSEWLDGIR L*

```

5 ORF105a and ORF105-1 show 93.8% identity in 291 aa overlap:

```

10 orf105a.pep      10      20      30      40      50      60
      MPTVRFTESVSKHDLDALEWAKASYGAECSWKTLYLNGPLGNLSPEWAEVRVKKQWEAG
orf105-1      10      20      30      40      50      60
      MPTVRFTESVSKQDLDALEWAKASYGAECSWKTLYLNGPLGNLSPEWAEVRVKKQWEAG

      70      80      90      100     110     120
orf105a.pep      CSESSDGIPLNADGWPMGRRLOHLARIWKEAGLLHGWDECFDLTDGGSNPLFALEAA
orf105-1      70      80      90      100     110     120
      CSESSDGIPLNADGWPMGRRLOHLALGWHCAGLLDGNWNECFDLTDGGSNPLFLERAA

      130     140     150     160     170     180
orf105a.pep      FRPFGLLSRAVHLNGLVESDGRWHFWIGRRSPHKAVDPKLDNTAAGGVSSGELPSETVC
orf105-1      130     140     150     160     170     180
      FRPFGLLSRAVHLNGLTESDGRWHFWIGRRSPHKAVDPKLDNTAAGGVSSGEMPSEAVC

      190     200     210     220     230     240
orf105a.pep      RESSEAGLIDKTLPLIRPVSQLHSLRPVSRGVHNEILYVFDVLPETFLPENQDGEVAG
orf105-1      190     200     210     220     230     240
      RESSEAGLIDKTLPLIRPVSQLHSLRPVSRGVHNEILYVFDVLPETFLPENQDGEVAG

      250     260     270     280     290
orf105a.pep      FEKMDIGGLLAAMLSGNMHDAQVLTDAFCRYGLIDAAHPLSEWLDGIRLX
orf105-1      250     260     270     280     290
      FEKMDIGGLLDAMLSGNMHDAQVLTDAFCRYGLIDAAHPLSEWLDGIRLX

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF105 shows 87.5% identity over a 312aa overlap with a predicted ORF (ORF105.ng) from *N. gonorrhoeae*:

```

40 orf105.pep      MVARRAHNPVKVGSNFXPATXFTQPRFNAEXVLXLPVSCFLFPAASVFCRIFLPAAISER 60
orf105ng      MVARRAHNPVKVGSNFPATKYQTPRFNAEGVLF-----FLFPAASVFCRIFLPAAISER 55

orf105.pep      QTAVCLRLQIQAVWLQSSALSSRKPTMPTVRFTESVSKQDLDALEWAKASYGAECSWK 120
orf105ng      QAIVCLRLQIQAVWLQSSALCSRKPAMPTVRFTESVSKQDLDALEWAKASYGAECSWK 115

orf105.pep      LYLNKPLGNLSPEWAEVRVKKQWEAGCESSESSDGIPLNADGWPMGRRLOHLALGWHCAGL 180
orf105ng      LYLNRLPLGNLSPEWAEIRIKKQWEAGCESSESSDGIPLNADGWPMGRRLOHLARTWNKAGL 175

orf105.pep      LDGWRNECFDLTDGGSNPLFTLERAXRPXGLLSRAVHLNGLTESDGRWHFWIGRRSPHK 240
orf105ng      LHGWNECFDLTDGGSNPLFTLERAAFRPFGLLIRAVHLNGLVESNGRWHFWIGRRSPHK 235

55 orf105.pep      AVDPNKLNTXAGGVSGGEMPSEAVCRESSEAGLIDKTLPLIRPVSQLHSLRPVSRGVH 300
orf105ng      AVDPKLDNIAGGVSGGEMPSEAVCRESSEAGLIDKTLPLIRPVSQLHSLRPVSRGVH 295

orf105.pep      NEILYVFDVLP 312
orf105ng      NEILYVFDVLPETFLPENQDGEVAGFEKMDIGGLLDAMLSGNMHDAQVLTDAFYRYG 355

```

A complete length ORF105ng nucleotide sequence <SEQ ID 415> was predicted to encode a protein having amino acid sequence <SEQ ID 416>:

1 MYARAHNPK VVGSNPAFAT KYQTPRNEAE GVLFLEFPAA SVFCRIFLPA
 51 AISRQDAVC LRLQICQWVI QSSALCSRK AMPTVRFTESV VSKQDLDALE
 101 ERAKASYGAE SCHWTLYLNR LPLGNLSPFW AERIKKDWEA CSESSDNGIF
 151 INADGWPMG GRLOHLARTW NKAGLHGRW NECFLTDGG GNPLFLTERA
 201 AFRPFLGILIR AVHINGLIVS NGRWHFWIGR RSPHKAVDPG KLDNIAGGGV
 251 SGGEMPSEAV CRESSEAGL DKTLFPLIRP VSRHLSLRPV SRGVHNEILY
 301 VFDVLPETTF LPENQDGEVA FEKMDIGGL LDAML SKNMH HDAQLVTLDA
 351 FYRYGLIDAA HPLSEWLDGI RL*

Further work revealed the complete nucleotide sequence <SEQ ID 417>:

1 ATGCCGACCG TCCGTTTAC CGAATCGGTC AGCAACAAG ACCTTGATGC
 51 CCTGTTTCGAG CGGGCAAAAG CAAGTTACGG TGCCGAAAGT TGCTGGAAAA
 101 CGCTGTATCT GAACCGTCTT CTTTGGGCA ATCTGCGCC GGAATGGGCT
 151 GAGCGCATCA AAAAAGACTG GGAGGCAGG TGCTCCGAGT CTTCAGACGG
 201 CATTTTTCTG AATGCGGAGC GCTGGCCGGA TATGGGCGGA CGCTTCGACG
 251 ACCTCGCCCG CACATGGAAC AAGCGGGGGC TGCTTCACGG ATGGCGCAAC
 301 GAGTGTTCG ACCTGACCGA CGGGCGGGCG AACCCCTTGT TCACGCTCGA
 351 ACGCGCCGCT TTCCGTCGGT TCGGACTACT CAGCGCGGCG GTCCATCTCA
 401 ACGGTTTGGT CGATTCGACG GCGCAGATGC ATTTTGGAT AGGCAGCGCG
 451 AGTCCGACCA AAGCATCGA CCOCGGCAG CTGCAACAATA TTGCGGGCGG
 501 CGGTGTTTCC GCGCGCGAAT TGCGCTCTGA AGCGCTGTGC CCGCAAGCA
 551 GCGAAGAAGC CGGTTTGGAT AAAACGCTGT TTCCGCTCAT CCGCCAGTA
 601 TCGCGGCTGC ACAGCCTTCG CCGGCTCAGC CAGGCTGTC ACAATGAAT
 651 CCGTGTATGT TGTGATGCGG TCCTGCCCGA AACCTTCTG CCGAAGATC
 701 AGGATGGCGA GGTAGCGGGT TTTGAAAAA TGGACATGG CGGCGTATTG
 751 GATGCCATGT TGTGAAAAA CATGATGCAC GACGGGCAAC TGGTTACGGT
 801 GGACGCGTTT TACCGTTACG GTCTGATTGA TCGCGCCCAT CCGTGTCCG
 851 AGTGGCTGGA CGGCATACGT TTATAG

This corresponds to the amino acid sequence <SEQ ID 418; ORF105ng-1>:

1 MPTVRFTESV SKQDLDALE RAKASYGAES CWKTLYLNR LPLGNLSPFWA
 51 ERIKKDWAE CSESSDGIPL NADGWPMG RLQHLARTWN KAGLHGRWN
 101 ECFDLTDGG NPLFLTERAA FRPFLLSRA VHLNGLVSN GRWHFWIGR
 151 SPHKAVDPG KLDNIAGGVS GGEMPSEAV CRESSEAGL KTLFPLIRPV
 201 SRLHSLRPVS RGVHNEILY VFDVLPETFL PENQDGEVAG FEKMDIGGLL
 251 DAML SKNMH HDAQLVTLDAF YRYGLIDAAH PLSEWLDGIR L*

35 ORG105ng-1 and ORF105-1 show 93.5% identity in 291 aa overlap:

		10	20	30	40	50	60
orf105-1.pep		MPTVRFTESVSKQDLDALEWAKASYGAESCWKTLYLNLPLGNLSPFWERUWKWWEAG					
orf105ng-1		MPTVRFTESVSKQDLDALEWAKASYGAESCWKTLYLNLPLGNLSPFWERUWKWWEAG					
40		10	20	30	40	50	60
		70	80	90	100	110	120
orf105-1.pep		CSESSDGIPLNADGWPMGRLQHLALGWHCAGLLDGRWNECFDLTDGGGNPLFLTERAA					
orf105ng-1		CSESSDGIPLNADGWPMGRLQHLALGWHCAGLLDGRWNECFDLTDGGGNPLFLTERAA					
		70	80	90	100	110	120
		130	140	150	160	170	180
orf105-1.pep		FRPFLLSRAVHLNGLTSESGRWHFWIGRRSPHKAVDPGKLDNIAGGVS					
orf105ng-1		FRPFLLSRAVHLNGLTSESGRWHFWIGRRSPHKAVDPGKLDNIAGGVS					
		130	140	150	160	170	180
		190	200	210	220	230	240
orf105-1.pep		RESSEAGLDKTLPIRIPVSRHLSLRPVSRGVHNEILYVFDVLPETFLPENQDGEVAG					
orf105ng-1		RESSEAGLDKTLPIRIPVSRHLSLRPVSRGVHNEILYVFDVLPETFLPENQDGEVAG					
		190	200	210	220	230	240
		250	260	270	280	290	
orf105-1.pep		FEKMDIGGLDAML SKNMHDAQLVTLDFAFCRYGLIDAAHPLSEWLDGIRLX					
orf105ng-1		FEKMDIGGLDAML SKNMHDAQLVTLDFAFCRYGLIDAAHPLSEWLDGIRLX					
		250	260	270	280	290	

Furthermore, ORF105ng-1 shows homology with a yeast enzyme:

```

sp|P41888|TNRJ3_SCHPO THIAMIN PYROPHOSPHOKINASE (TFK) (THIAMIN KINASE)
>gi|1076928|p|f|S52350 thiamin pyrophosphokinase (EC 2.7.6.2) - fission yeast
(Schizosaccharomyces pombe) >gi|666111 (X84417) thiamin pyrophosphokinase
[Schizosaccharomyces pombe] >gi|2330852|gnl|P|D|e334056 (Z98533) thiamin
pyrophosphokinase [Schizosaccharomyces pombe] Length = 569
Score = 105 bits (259), Expect = 4e-22
Identities = 64/192 (33%), Positives = 94/192 (48%), Gaps = 3/192 (1%)

Query: 268 NKAGLHLGWRNECFDLTDGGNGLFTFLERAAPFPGLLSRAVHGLNGLVESNGRW--HEWI 441
      N G+ WRNE + + + Pt +ER F FG LS VH + + W+
Sbjct: 96 NTFGIADQWRNELYTVYGSKKPVLAVERGGFWLFGFLSTGVHCTMYIPATKEHPLRIWV 155

Query: 442 GRRSPHKAVDPKGLDNIAGGGVSGGEMPSEAVCRSESEAGLDKTLFPLIRPVSRHLHSLR 621
      RRSF K P LDN GG++ G+ + +E SEA LD + LI P + + +
Sbjct: 156 PRRSPTKQTWPNYLDSNVAGGIARGDSVIGTMIKEFSEANLDVSSMNLIPCGTYSYIK 214

Query: 622 PVSRG-VHNEILYVDAVLPEFLPENQDGEVAGFEKMDIGGLDAMLKSNMMDAQLVT 798
      R + E+ YVFD + + +P DGEVAGF + + +L + K+ + LV
Sbjct: 215 MEKRHWIQPELQYVFDLPVDDLVIPIRINDGEVAGFSLLPLNQLVHELELKSFKPNCAVL 274

Query: 799 LDAFYRYGLIDAAHP 843
      LD R+G+I HP
Sbjct: 275 LDFLIRHGIITPQHP 289

```

- 25 Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 49

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 419>:

```

1 ATGAATAGAC CCAAGCAACC CTTCITCCGT CCCGAAGTCG CGGTGCCCCG
51 CCAAACCAAGC CTGACGGGTA AAGTGATCTG GACACGACCG TTGTCACTTTT
101 CCCTATGGAG GACATTTCGA TCGATATCTG CGTATTGAT TATCCTGTTT
151 TTGATATTGT GTAACATATAC GCGAAGAACA ACAGTTGGAGG GACAAATTTT
351 ACCTGCATCG GCGGTAATCA GGTGTATGC ACCGATACG TgKACAAATTA
251 CAGCGAAATT CGTGGAAGAT GgmsAAAAAG TTAAGGCTGG CGACAAGTGA
301 TTTGCGCTTT CGACCTCAGC TTTTCGGCGCA GGAGGTAGCG TGCAGCAGCA
351 GTTGAAACAG GAGGCAGTTT TGAAGAAACG GTTGGCAGAA CAGGAACTGG
401 GTCGCTCGAA CGTGAACAC GGAATGAAA CGCGCagCct TAAAGCAACT
451 GTCGAACGTT TGGAAACCA GGAATCCAT ATTTCGCAAC AGATAGACGG
501 TCAGAAAAGG CGCATTAGAC TTGCGGAAGA AATGTTGCAG AAATATCGTT
551 TCCTATCCGC .CAATGA

```

This corresponds to the amino acid sequence <SEQ ID 420; ORF107>:

```

1 MNRPKQFFFR PEVAVARQTS LTGKVILTRP LSFSLWTTFA SISALLIILF
51 LIFGNYTRKT TVEGQILPAS GVIRVYAPDT XTITAKFVED GKXKAGDKL
101 FALSTSRFGA GGSVQQQLKT EAVLKLTLAE QELGRLLKLIH GNETRSLKAT
151 VERLENQELH ISQQIDGQKR RIRLAEMQLQ KYRFLSXQ*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

- 50 ORF107 shows 97.8% identity over a 186aa overlap with an ORF (ORF107a) from strain A of *N.meningitidis*:

10 20 30 40 50 60

	orf107.pep	MNRPKQFFRPEVAVARQTS	LTGKVLTRPLS	SFSLWTT	FASISALLIILFLIFGN	YTRKT	
	orf107a	MNRPKQFFRPEVAVARQTS	LTGKVLTRPLS	SFSLWTT	FASISALLIILFLIFGN	YTRKT	
5		10	20	30	40	50	60
	orf107.pep	TVEGQILPASGVIRVYAPD	XTITAKFVEDG	GKVKAGDKL	FALSTSRFGAGGS	VQQQLKT	
	orf107a	TVEGQILPASGVIRVYAPD	TGTITAKFVEDG	EKVKAGDKL	FALSTSRFGAGGS	VQQQLKT	
10		70	80	90	100	110	120
	orf107.pep	EAVLKKT	LAEQELGR	KLHGNETR	SLKATVERLENQEL	HISQQIDG	KRRIRLAEEM
	orf107a	EAVLKKT	LAEQELGR	KLHGNETR	SLKATVERLENQEL	HISQQIDG	KRRIRLAEEM
15		130	140	150	160	170	180
	orf107.pep	EAVLKKT	LAEQELGR	KLHGNETR	SLKATVERLENQEL	HISQQIDG	KRRIRLAEEM
	orf107a	EAVLKKT	LAEQELGR	KLHGNETR	SLKATVERLENQEL	HISQQIDG	KRRIRLAEEM
20		190	200	210	220	230	
	orf107.pep	KYRFLSXQ					
	orf107a	KYRFLS	SANDAVPKQEM	MNVKAE	LLEQKAKL	DAYRRE	EVGLLQEI

The complete length ORF107a nucleotide sequence <SEQ ID 421> is:

25	1	ATGAATAGAC	CCAAGCAACC	NTTCTCCGT	CCGGAAGTCG	CGSTTGCCCG	
	51	CCAAACACCG	CTGACGGGTA	AAGTGATTCT	GACACGACCG	TTGTCATTTT	
	101	CCCTATGAC	GACATTGCA	TGCGATCTG	CGTTATTGAT	TATCCTGTTT	
	151	TTGATATTG	GTAACATATC	GCGAAGACA	ACAGTGGAGG	GACAAATTTT	
	201	ACCTGCATCG	GCGCTAATCA	GGGTGTATGC	ACCGGATACG	GGGCAATTA	
	251	CNCGAARAT	CNTGGAAGAT	GGAGAAAGG	TTAAGGCTGG	CGACAAGCTA	
30	301	TTTGGCTTT	CGACCTCAGC	TTTCCGGCGA	GGAGATAGCG	TGCAGCAGCA	
	351	GTTGAARACG	GAGGCAGTTT	TGAAGAAAC	GTTGGCAGAA	CAGGAAGCTG	
	401	GTCGTCGAA	GCTGATACAC	GGGAATGAAA	CGCGCAGCTT	TAAAGCAACT	
	451	GTGGAACGTT	TGCGAARACCA	GGAACTCCAT	ATTTCGCAAC	AGATAGACGG	
	501	TCGGAAGAG	GCATATGAC	TTCCGGGACA	AATGTCGAG	AATATCGTT	
35	551	TCCTATCGCG	CAATGATCCA	GTCCCAAAAC	AAGAAATGAT	GAATGTCAG	
	601	GCACAGCTTT	TAGAGCAGAA	AGCCAAACTT	GATGCTTACC	GCCGAGAGAA	
	651	AGTCGGGCTG	CTTCAGGAAA	TCCGACGCA	GAATCTGACA	TTGNNAGCC	
	701	TCCCCAAGC	GGCATGA				

This encodes a protein having amino acid sequence <SEQ ID 422>:

40	1	MNRPKQFFR	PEVAVARQTS	LTGKVLTRP	LSFSLWTTFA	SISALLIILF	
	51	LIFGN	YTRKT	TVEGQILPAS	GVIRVYAPDT	GTITAKFXED	GEKVKAGDKL
	101	FALSTSRFGA	GDSVQQQLKT	EAVLKKT	LAEQELGR	KLHGNETR	SLKAT
	151	VERLENQELH	ISQQIDGQKR	RIRLAEEM	LQ	KYRFLS	SANDA VPKQEM
45	201	AELLEQAKL	DAYRRE	EVGLLQEI	RNTL	LXSLPQAA*	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF107 shows 95.7% identity over a 188aa overlap with a predicted ORF (ORF107.ng) from *N. gonorrhoeae*:

50	orf107.pep	MNRPKQFFRPEVAVARQTS	LTGKVLTRPLS	SFSLWTT	FASISALLIILFLIFGN	YTRKT	60
	orf107ng	MNRPKQFFRPEVAIARQTS	LTGKVLTRPLS	SFSLWTT	FASISALLIILFLIFGN	YTRKT	60
	orf107.pep	TVEGQILPASGVIRVYAPD	XTITAKFVEDG	GKVKAGDKL	FALSTSRFGAGGS	VQQQLKT	120
55	orf107ng	TMEGQILPASGVIRVYAPD	TGTITAKFVEDG	EKVKAGDKL	FALSTSRFGAGGS	VQQQLKT	120
	orf107.pep	EAVLKKT	LAEQELGR	KLHGNETR	SLKATVERLENQEL	HISQQIDG	KRRIRLAEEM
	orf107ng	EAVLKKT	LAEQELGR	KLHGNETR	SLKATVERLENQEL	HISQQIDG	KRRIRLAEEM
60	orf107.pep	KYRFLSXQ	188				
	orf107ng	KYRFLSAQ	188				

The complete length ORF107ng nucleotide sequence <SEQ ID 423> is predicted to encode a protein having amino acid sequence <SEQ ID 424>:

```

1  MNRPKQPFER PEVAIARQTS LTGKVILTRP LSFSLWTTFA SISALLIILE
5  51  LIFGNYTRKT TMEGQILPAS GVRVYAPDT GTITAKFVED GEKVAGDKL
101 101  FALSTSRFGA GGSVOOQLTK EAVLKTLAE QELGRKLIIH ENETRSLKAT
151 151  VERLENOKLH ISQIDGQKR RIRLAEMELR KYRFLSAQ*

```

Based on the presence of a putative ransmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

10 Example 50

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 425>:

```

1  ATGCTGAATA CTTTTTTTGC CGTATTGGGC GGCTGCGTGC TGCT.TTGGC
51 51  GTGCGGCAAA TCCTGTAATA CGCGGTAACA CGCGCAAAAC GCGGTACAAA
101 101  CGCGGCGGAA ACGCGTTTTC AAGTCATAT ATATCGACAA TACGGCGATT
151 151  CGCGGTTTGG ATTTGGGACA AAGCAGCGAA GCGCAAAACA ACGACGCGAA
201 201  AAAACAATAT AGTTATCCGA TTAAAGCCTT GCGGAAACAA AATGTTATCC
251 251  GACTGATCGG CAAGCATCCC GCGCATCTGG AAGCGCTCAG CGGCAATGTT
301 301  ATGGAACCGG ATGATAAGGA CAGTCGCGCA GGTGGGCGAG AAAACGGCGT
20 351  GTGCCATACC TTGTTTGCCA AACTGCTGGG CAATATCGCC GAAGACGGCG
401 401  GCAAACTGAC GGATTACCTA GTTTCGCGAT CGCGCTGCGA ACCCTATCAG
451 451  GCAGGCAAAA CGCGCTATCG CGCGTGCAG AACGGACGCT ATGTGCTGGA
501 501  AATCGACAGC GAAGGGGCGT TTTATTTCCG CGCGCGCAT TATTGA

```

This corresponds to the amino acid sequence <SEQ ID 426; ORF108>:

```

25 1  MLNTEFAVLG GCLLLXPCGK SVNTAVQPQN AVQSAKPFV KVIYIDNTAI
51 51  AGLDLGQSSE GKTNDGKKQI SYPIKGLPEQ NVIRLIGKHP GDLEAVSGKC
101 101  METDDKDSFA GWAENGVCHT LFAKLVGNIA EDGKGLTDYL VSHAALQPYQ
151 151  AGKSGYAAVQ NGRYVLEIDS EGAFYFRRRH Y*

```

Further work revealed the following DNA sequence <SEQ ID 427>:

```

30 1  ATGCTGAAAA CATCTTTTGC CGTATTGGGC GGCTGCGTGC TGCTTGCGCG
51 51  CTGCGGCAAA TCCTGTAATA CGCGGTAACA CGCGCAAAAC GCGGTACAAA
101 101  CGCGGCGGAA ACGCGTTTTC AAGTCATAT ATATCGACAA TACGGCGATT
151 151  CGCGGTTTGG ATTTGGGACA AAGCAGCGAA GCGCAAAACA ACGACGCGAA
201 201  AAAACAATAT AGTTATCCGA TTAAAGCCTT GCGGAAACAA AATGTTATCC
35 251  GACTGATCGG CAAGCATCCC GCGCATCTGG AAGCGCTCAG CGGCAATGTT
301 301  ATGGAACCGG ATGATAAGGA CAGTCGCGCA GGTGGGCGAG AAAACGGCGT
351 351  GTGCCATACC TTGTTTGCCA AACTGCTGGG CAATATCGCC GAAGACGGCG
401 401  GCAAACTGAC GGATTACCTA GTTTCGCGAT CGCGCTGCGA ACCCTATCAG
451 451  GCAGGCAAAA CGCGCTATCG CGCGTGCAG AACGGACGCT ATGTGCTGGA
40 501  AATCGACAGC GAAGGGGCGT TTTATTTCCG CGCGCGCAT TATTGA

```

This corresponds to the amino acid sequence <SEQ ID 428; ORF108-1>:

```

45 1  MLKTSFAVLG GCLLLAACGK SENTAEQPN AVQSAKPFV KVIYIDNTAI
51 51  AGLDLGQSSE GKTNDGKKQI SYPIKGLPEQ NVIRLIGKHP GDLEAVSGKC
101 101  METDDKDSFA GWAENGVCHT LFAKLVGNIA EDGKGLTDYL VSHAALQPYQ
151 151  AGKSGYAAVQ NGRYVLEIDS EGAFYFRRRH Y*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF108 shows 88.4% identity over a 181aa overlap with a predicted ORF (ORF108.ng) from *N.gonorrhoeae*:

5	orf108.pep	MLNTFFAVLGGCLLXLPCKGSVNTAVOPONAVOSAPKPFVKVIYDNTAIALGDLGQSSE	60
	orf108.ng	MLKIPFAVLGGCLLLAACGKSENTAEQPNAAQSAKPFVKVIYDNTAIALGDLGQSSE	60
	orf108.pep	GKTNDGKKQISYPIKGLPEQNVIRLIGKH PGDLEAVSGKCMETDDKDSFAGWAENGVCHT	120
10	orf108.ng	GKTNDGKKQISYPIKGLPEQNAVRLTGKHPNDLEAVVGKCMETDGKDAPSGWAENGVCHT	120
	orf108.pep	LFAKLVGNIAEDGGKLT DYLVSHAA LQPYQAGKSGYAAVQNGRYVLEIDSEGAIFYRRRHY	181
15	orf108.ng	LFAKLVGNIAEDGGKLT DYLVSHAA LQPYQAGKSGYAAVQNGRYVLEIDSEGAIFYRRRHY	181

ORF108-1 shows 92.3% identity with ORF108ng over the same 181 aa overlap:

	orf108-1.pep	MLKTSFAVLGGCLLLAACGKSENTAEQPNAAQSAKPFVKVIYDNTAIALGDLGQSSE	60
	orf108ng-1	MLKIPFAVLGGCLLLAACGKSENTAEQPNAAQSAKPFVKVIYDNTAIALGDLGQSSE	60
20	orf108-1.pep	GKTNDGKKQISYPIKGLPEQNVIRLIGKH PGDLEAVSGKCMETDDKDSFAGWAENGVCHT	120
	orf108ng-1	GKTNDGKKQISYPIKGLPEQNAVRLTGKHPNDLEAVVGKCMETDGKDAPSGWAENGVCHT	120
25	orf108-1.pep	LFAKLVGNIAEDGGKLT DYLVSHAA LQPYQAGKSGYAAVQNGRYVLEIDSEGAIFYRRRHY	181
	orf108ng-1	LFAKLVGNIAEDGGKLT DYLVSHAA LQPYQAGKSGYAAVQNGRYVLEIDSEGAIFYRRRHY	181

The complete length ORF108ng nucleotide sequence <SEQ ID 429> is:

30	1	ATGCTGAAAa	tacctTTTGC	CGTGTgtggc	ggCtgccctgc	TGCTTGCCGC
	51	CTGCGGCAAA	TCGAAAATa	cggcggaacA	GCCGCAAAAT	gcggCACAAA
	101	GCGGCGCAAA	ACCGTTTTC	AAAGTCAAT	ACATCGACAA	TACGGCGATT
	151	GCGCGTTTGG	CTTTGGGACA	AAATAGCGAA	GGCAAAACCA	acggcgcGAA
	201	AAAACAAATC	AGTTATccgA	TTAAGGCTT	GCCGGAACAA	Aacgcgctcc
	251	gGCTGACCGG	AAAGCATCC	AACGACTTG	AagccgtcgT	CGSCAAATGT
35	301	ATGGAACCCG	ACGGAAGGA	CGCGCCTCG	GCGTGGCGGG	AAACCGCGGT
	351	GTGCCATACC	TTGTTTGCCA	AACTGGTGGG	CAATATCGCC	GAAGACGGCG
	401	GCAAACTGAC	TGATTACCTG	ATTTCGCAAT	CCGCCCTGCA	ACCCATATCAG
	451	GCAGGCAAAA	GCGGCTATCG	CGCCGTGCAG	AACGACGCT	ATGTGCTGGA
	501	AATCGACAGC	GagggGGCGT	TTTATttccg	cgcgcgcat	tattgA

40 This encodes a protein having amino acid sequence <SEQ ID 430>:

```

1  MLKIPFAVLG  GCLLLAACGK  SENTAEQPN  AAQSAKPFV  KVKYIDNTAI
51  AGLALGQSSE  GKTNDGKKQI  SYPIKGLPEQ  NAVRLTGKHP  NDLEAVVGKC
101  METDGKDAPS  GWAENGVCHT  LFAKLVGNIA  EDGKGLTDYL  ISHSALQPYQ
151  AGKSGYAAVQ  NGRYVLEIDS  EGAFYFRRRH  Y*
```

45 Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) and a putative ATP/GTP-binding site motif A (P-loop, double-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

50 Example 51

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 431>:

```

1  ATGGAAGATT  TATATATAAT  ACTCGCTTTG  GGTTCGGTTC  CGATGATTGC
51  CGGATTTATC  GATCGGATTG  cgggCGGGGG  TGGTTGATT  ACGCTGCCCG
101  CACTCTTGTT  GGCAGGTATT  CTTCCCGTGT  CGGCAATTGC  CACCACACAG
151  CTGCAAGCAG  CCGCTGCTAC  GTTTTCAGCT  ACGGTTCTCT  TTGACGCAAA
201  AGGTTTGATT  GATTGGAAGA  AAGGTCCTCC  GATTGCGCGA  GCATCGTTTG
5  251  TAGCGCGCGT  GGC CGGTGCA  TTATCGGTCA  GCTTGCGTTC  CAAAGATATT
301  CTgCTgCGCG  TCGTGC CGGT  TTTGTTGATA  TTTGTCGCAC  TGTATTTTGT
351  GTTTTCGCCCC  AAGCTCGACG  GCAGTAAGGA  AGGCAAAGCC  AGAATGCTCT
401  TTTTTCCTGTT  cGGGCTGACG  GTGCG. ACCG  CTTTGGGGTT  TTTACGACGG
10  451  TGTGTTGCGA  CCGGGTGTCG  GCTCGTTTTT  TCTGATTGCC  TTTATTTGTT
501  TGCTCGGCTG  CAAGCTGTTG  AACGCGATGT  CTTACACCAA  ATTGGCGAAC
551  GTTGCGTGCA  ATCTTGTTTC  GCTATCGGTA  TTCCTGCTCG  ACGGTTTCAT
601  TATTTTCCCG  ATTGCGGCAA  CGATGCGGCT  CGGTGCGTTC  GTCGGTGC GA
651  ATTTAGGTGC  GAGATTGGCC  GTAcgctTCG  GTTCGAAGCT  GATTAA

```

15 This corresponds to the amino acid sequence <SEQ ID 432; ORF109>:

```

1  MEDLYIILAL  GLVAMIAGFI  DAIAGGGGLI  TLPALLLAGI  PPVSAIATNK
51  LQAAAAATFSA  TVSFARKGLI  DWKKGPIAA  ASFVGGVAGA  LSVSLVSKDI
101  LLAVVPVLLI  FVALYFVFSF  KLDGSKGKA  RMSFFLFLGT  VXTAFGLRFR
151  CVRTGCRILVF  SDCLYCFARL  QAVRDVLHQ  IGERCLQSWF  AIGIPARFDF
20  201  YFFCGNDGG  RCVCRCFEFC  EICRTLRFEA  D*

```

Further work revealed the following DNA sequence <SEQ ID 433>:

```

1  ATGGAAGATT  TATATATAAT  ACTCGCTTTG  GGTTCGGTTC  CGATGATTGC
51  CGGATTTATC  GATCGGATTG  CCGGCGGGGG  TGGTTGATT  ACGCTGCCCG
25  101  CACTCTTGTT  GGCAGGTATT  CTTCCCGTGT  CGGCAATTGC  CACCACACAG
151  CTGCAAGCAG  CCGCTGCTAC  GTTTTCAGCT  ACGGTTCTCT  TTGACGCAAA
201  AGGTTTGATT  GATTGGAAGA  AAGGTCCTCC  GATTGCGCGA  GCATCGTTTG
251  TAGCGCGCGT  GGC CGGTGCA  TTATCGGTCA  GCTTGCGTTC  CAAAGATATT
301  CTGCTGCGGG  TCGTGC CGGT  TTTGTTGATA  TTTGTCGCAC  TGTATTTTGT
351  GTTTTCGCCCC  AAGCTCGACG  GCAGTAAGGA  AGGCAAAGCC  AGAATGCTCT
30  401  TTTTTCCTGTT  CCGGCTGACG  GTGCGACCGC  TTTTGGGGTT  TTACACGGGT
451  GTGTTGCGAC  CCGGTGTGCG  CTCGTTTTTT  CTGATTGCCT  TTTATTTGTT
501  GCTCGGCTGC  AAGCTGTTGA  ACGCGATGTC  TTACACCAA  TTGCGGAAGC
551  TTGCGTGC AA  TCTTGGTTCG  CTATCGGTAT  TCTGCTGCGA  CGGTCGATT
35  601  ATTTTCGCGA  TTGCGGCAAC  GATGCGGGTC  GGTGCGTTG  TCGTGCAGAA
651  TTTAGGTGCG  AGATTGCGCG  TCGCTTCGG  TTCGAAGCTG  ATTAAGCCGC
701  TGCTGATTGT  CATCAGCATT  TCGATTGGCT  TGAATTTGTT  GATAGACGAG
751  AGAAATCGCG  TGTATCAGAT  GATGTTTTCG  ATGTTTTTAA

```

This corresponds to the amino acid sequence <SEQ ID 434; ORF109-1>:

```

1  MEDLYIILAL  GLVAMIAGFI  DAIAGGGGLI  TLPALLLAGI  PPVSAIATNK
40  51  LQAAAAATFSA  TVSFARKGLI  DWKKGPIAA  ASFVGGVAGA  LSVSLVSKDI
101  LLAVVPVLLI  FVALYFVFSF  KLDGSKGKA  RMSFFLFLGT  VAPLLGFYDG
151  VRFPGVGSFF  LIAFTVLLGC  KLLNAMS YTK  LANVACNLGS  LSVFLLHGS I
201  IFPIAATMAV  GA*FVGNLGA  RFAVRGSKL  IKPLLVISVI  SMAVKLLIDE
251  RNFLYQIVS  MF*

```

45 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF109 shows 95.9% identity over a 147aa overlap with an ORF (ORF109a) from strain A of *N.*

meningitidis:

```

50  orf109.pep  MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKIQLQAAAAATFSA
orf109a      MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKIQLQAAAAATFSA
              10      20      30      40      50      60
55  orf109.pep  TVSFARKGLIDWKKGPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSF
orf109a      TVSFARKGLIDWKKGPIAAASFAGGVVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSF
              70      80      90      100     110     120

```

-263-

		130	140	150	160	170	180
orf109.pep		KLDGSEKKGARMSFFFLGFLTVXTAFGLRRCVRTGCLVFSDCLYCFARLQAVRDVLHQ					
5	orf109a						
		130	140	150	160	170	180
		KLDGSEKKGARMSFFFLGFLTVAPLLGFYDGVFGPGVGSFFFLIAFIVLLGCKLLNAMS					

The complete length ORF109a nucleotide sequence <SEQ ID 435> is:

	1	ATGGAAGATT	TATACATAAT	ACTCGCTTTG	GGTTTGGTTG	CGATGATTGC
	51	CGGATTTATC	GATGCGATTG	CGGGTGGGGG	TGGTTTGATT	ACGCTGCCTG
10	101	CACCTCTGTT	GGCAGSTATT	CCCTCGGTGT	CGGCAATTGC	CACCAACAAG
	151	CTGCAAGCAG	CGCGTCTAC	GTTTTCGGCT	ACGGTTCTCT	TTGCAACGCA
	201	AGGTTTGATT	GATTGGAAGA	AAGGTCCTCC	GATTGCGGCA	GCATCGTTTG
	251	CAGGCGGCGT	GTCGCTGCA	TTATCGGTCA	GCTTGGTTTC	CAAGAATATT
	301	CTGCTGGCGG	TCGTGCGGTT	TTTGTGATA	TTTGTGCGCG	TGTAATTTGT
15	351	GTTTTCGCC	AAGCTCGAAG	GCAGTAAGGA	AGGCAAGGCC	AGAATGTCTT
	401	TTTTCTGTT	CGGCTGACG	GTGCAACAC	TTTTGGGTTT	TTACGACGGT
	451	GTTCTCGGAC	CGGGTGTCCG	CTCGTTTTT	CTGATTCGCT	TTATTTGTTT
	501	GTCGGCTGCG	ANGCTGTTGA	ACGCGATGTC	TTACACCAAA	TTGCGCAAGC
	551	TTGCGTGCAG	TCTTGCTTCG	CTATCGGTAT	TCCTGCTGCA	CGGTTGCAAT
20	601	ATTTCCTCGA	TTGCGGCAAC	GATGGCGGCT	GGTGGGTTTG	TCGCTGCGAA
	651	TTTAGGTGCG	AGATTGTCGG	TCGCTTCGCG	TTGCAAGCTG	ATTAAGCCCG
	701	TGCTGATTGT	CATCAGCATT	TCGATGGCTG	TGAATATTGT	GATAGACGAG
	751	AGAAATCCCG	TGTATCAGAT	GATTGTTTGG	ATGTTTAA	

This encodes a protein having amino acid sequence <SEQ ID 436>:

	1	MEDLYIILALGLVAMIAGFI	DAIAGGGGLI	TLFALLLAGI	PFVSAIATNK
	51	LQAAATFSA	TVSFARKGLI	DWKKGLPIAA	ASFAGGVVGA
	101	LLAVVPEVLLI	FVALYFVFSP	KLDGSEKGRK	RMSFFFLGTL
	151	VFGPGVGSFF	LFAFIVLLGC	KLLNAMS	YTK LSVFLLHGS
30	201	IFPIAATNAV	GAFVGANLGA	RFVRFSGSL	IKPELLIVISI
	251	RNPLYQMIVS	NF*	SMVAKLLID	

ORF109a and ORF109-1 show 99.2% identity in 262 aa overlap:

		10	20	30	40	50	60
	orf109a.pep	MEDLYIILALGLVAMIAGFIDA	IAGGGGLITL	PALLLAGI	PFVSAIATNK	LQAAATFSA	
35	orf109-1	MEDLYIILALGLVAMIAGFIDA	IAGGGGLITL	PALLLAGI	PFVSAIATNK	LQAAATFSA	
		10	20	30	40	50	60
	orf109a.pep	TVSFARKGLIDWKKGLPIAA	ASFAGGVGAL	SVSLVSKD	ILLAVVPEVLLI	FVALYFVFSP	
40	orf109-1	TVSFARKGLIDWKKGLPIAA	ASFAGGVGAL	SVSLVSKD	ILLAVVPEVLLI	FVALYFVFSP	
		70	80	90	100	110	120
	orf109a.pep	TVSFARKGLIDWKKGLPIAA	ASFAGGVGAL	SVSLVSKD	ILLAVVPEVLLI	FVALYFVFSP	
	orf109-1	TVSFARKGLIDWKKGLPIAA	ASFAGGVGAL	SVSLVSKD	ILLAVVPEVLLI	FVALYFVFSP	
		130	140	150	160	170	180
45	orf109a.pep	KLDGSEKKGARMSFFFLGFLTV	APLLGFYDGV	FGPGVGSFF	LIAFIVLLGCKLLNAMS	YTK	
	orf109-1	KLDGSEKKGARMSFFFLGFLTV	APLLGFYDGV	FGPGVGSFF	LIAFIVLLGCKLLNAMS	YTK	
		130	140	150	160	170	180
50	orf109a.pep	LANVACNLGSLSVFLHGS	IIFPIAATNAV	GAFVGANL	GARFAVRFGSK	LKPELLIVISI	
	orf109-1	LANVACNLGSLSVFLHGS	IIFPIAATNAV	GAFVGANL	GARFAVRFGSK	LKPELLIVISI	
		190	200	210	220	230	240
55	orf109a.pep	SMVAKLLIDERNPLYQMIVS	SMFX				
	orf109-1	SMVAKLLIDERNPLYQMIVS	SMFX				
60		250	260				

Homology with a predicted ORF from *N. gonorrhoeae*

ORF109 shows 98.3% identity over a 231aa overlap with a predicted ORF (ORF109.ng) from *N.*

gonorrhoeae:

5	orf109.pep	MEDLYIILALGLVMIAGFIDAIAGGGGLITLPALLLAGIPVSAIATNKLQAAATFSA	60
	orf109ng	MEDLYIILALGLVMIAGFIDAIAGGGGLITLPALLLAGIPVSAIATNKLQAAATFSA	60
10	orf109.pep	TVSFARKGLIDWKKGPIAAASFVGGVAGALSVLSVSKDILLAVVPVLLIFVALYVFVFS	120
	orf109ng	TVSFARKGLIDWKKGPIAAASFVGGVAGALSVLSVSKDILLAVVPVLLIFVALYVFVFS	120
15	orf109.pep	KLDGSKGEGKARMSFFLFGLTVXATFGFLRRCVRTGCRLVFSDCLYCFARLQAVRDVLHQ	180
	orf109ng	KLDGSKGEGKARMSFFLFGLTVATFGFLRRCVRTGCRLVFSDCLYCFARLQAVRDVLHQ	180
	orf109.pep	IGERCLQSWFAIGIPAAARFDYFPDGNDDGRCVCRCEFRCEICRTLRFAD	231
	orf109ng	IGERCLQSWFAIGIPAAARFDYFPDGNDDGRCVCRCEFRCEICRPLRFAD	231

An ORF109ng nucleotide sequence <SEQ ID 437> was predicted to encode a protein having amino

20 acid sequence <SEQ ID 438>:

25	1	MEDLYIILAL GLVMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
	51	LQAAATFSA TVSFARKGLI DWKKGPIAA ASFAGGVGA LSVLSVKDI
	101	LLAVVPVLLI FVALYVFVSP KLDGSKGEGA RMSFFLFGLT VATAFGLRR
	151	CVRTGCRLVF SDCLYCFARL QAVRDVLHQ IGERCLQSWF AIGIPAAARFD
	201	YFPDGNDDG RCVCRCEFR EICRPLRFEA D*

Further work revealed the following gonococcal DNA sequence <SEQ ID 439>:

30	1	ATGGAAGATT TATACATAAT ACTCGCTTTG GGTTTGGTTG CGATGATCGC
	51	CGGATTATAT GATGCGATTG CGGGCGGGGG TGGTTTGATT ACGCTGCCTG
	101	CACCTCTTGT GGCAGGTATT COTCCCGTGT CGGCAATTGC CACCAACAAG
	151	CTGCAAGCAG CCGCTGCTAC GTTTTCGGCT ACGGTTTCTT TTGCACGCAA
	201	AGGTTTGATT GATTGGAAGA AAGGTCCTCC GATTGCGCGA GCATCGTTTG
	251	CAGCGCGCGT GTCGCTGCA TTAGCTGTCA GCTTGGTTTC CAAAGATATT
	301	TTGCTGCGCG TCGTCCGCGT TTGTGTGATA TTGTGCGCGC GTATTATTGT
	351	GTITTCGCGC AAGCTGCAAG GCATGAGGA AGCAAGCC AGAATCTCTT
35	401	TTTTCTATT CGGCTGAGG GTTGACCCGC TTTTGGGTTT TTACGACGCT
	451	GTGTTCCGAC CGGCTGCGG CTCGTTTTTT CTGATTGCCT TTATTGTTTT
	501	GCTCGGCTGC AAGCTGTGA ACGGATGTC TTACACCAA TTGGCGAAGC
	551	TTGCTTGCAA TCTTGGTTGC CTATCGGTAT TCCTGCTGCA CGGTTGATT
40	601	ATTTTCCGCA TTGTGGCAAC GATGCGGTC GGTGGGTTG TCGGTGCGAA
	651	TTTAGTGGCG AGATTGCGG TCCGCTTCGG TTCAAGCTGC ATTAAGCCGC
	701	TGCTGATTGT CATCAGCATT TCGATTGCTG TGAATTTGTT GATGACCGAG
	751	AGAAATCCGC TGTATCAGAT GATTGTTTGC ATGTTTAA

This corresponds to the amino acid sequence <SEQ ID 440; ORF109ng-1>:

45	1	MEDLYIILAL GLVMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
	51	LQAAATFSA TVSFARKGLI DWKKGPIAA ASFAGGVGA LSVLSVKDI
	101	LLAVVPVLLI FVALYVFVSP KLDGSKGEGA RMSFFLFGLT VAPLGLGYDG
	151	VFGPGVGSFF LIAFIVLLGC KLINAMSYTK LANVACNLGS LSVFLHGSII
	201	IFP1VATMAV GAFVGNLGA RFAVRFGSKL IKPLIVISII SHAVKLLIDE
	251	RNPLYQMIVS MF*

50 ORF109ng-1 and ORF109-1 show 98.9% identity in 262 aa overlap:

55		10	20	30	40	50	60
	orf109ng-1.pep	MEDLYIILALGLVMIAGFIDAIAGGGGLITLPALLLAGIPVSAIATNKLQAAATFSA					
	orf109-1	MEDLYIILALGLVMIAGFIDAIAGGGGLITLPALLLAGIPVSAIATNKLQAAATFSA					
		10	20	30	40	50	60
		70	80	90	100	110	120
	orf109ng-1.pep	TVSFARKGLIDWKKGPIAAASFAGGVGALSVLSVSKDILLAVVPVLLIFVALYVFVFS					

[illegible]

In addition, ORF109ng-1 shows homology to a hypothetical *Pseudomonas* protein:

```

25  spl229942|YCB9 PSDEE HYPOTHETICAL 27.4 KD PROTEIN IN COBO 3'REGION (ORF9)
>gi|949842|pir|I38164 hypothetical protein 9 - Pseudomonas sp >gi|551929
(M52866) ORF9 [Pseudomonas denitrificans] Length = 261
Score = 175 bits (439), Expect = 3e-43
Identities = 83/214 (38%), Positives = 131/214 (60%), Gaps = 1/214 (0%)

30  Query: 41  PPSAIAITNKLQXXXXXXXXXXXXXKRLGGLIDKKGLPIXXXXXXXXXXXXXXXXXXXX 100
PP+ + TNKLQ R+G + K LP+ D+
Sbjct: 43  PPLQTLFTGNKLQGLFGSGSATLSYARRGHVNLKQLPALMSAAGAVLGALLATIVPGDV 102

Query: 101  LLAVVPVLLILFVALYFVPSFKLQSGKEKGARMSFFLFGTLVAPLLGFYDVGVPGVGSFF 160
L A++P LLL +ALYF P + G + R++ F+L+ PL+DGFVGVEGPG GSF
35  Sbjct: 103  LKAILPELLIALIALYFGLKPMN-GDVQDHRSRVTFPVFTLITLVELIGFYDVGVEGPGTSFF 161

Query: 161  LIAIFLVLGCKLLNAMSYSKLANVNCNLSGSLVFLHSGSIIFVIVATMVGAVFGANLGA 220
+ F+G+ GL A +N+ VEL M +G F+G+ +
Sbjct: 162  MLGFVTLTAGVFGKLAATKATHLFLNFGSNVGAVGVFLFPGVWLKVGVLMLGLQFLGAQVS 221

40  Query: 221  RFNRVFGSKLILKPLLVLSISMVAKLIDERNPL 254
R+A+ G+K+IKPL+++SI++A++LL D +PL
Sbjct: 222  RYMAKAGKAIKPLLVIVISIAIAIRLADHPDL 255

```

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 52

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 441>:

50	1	..CTGCTAGGG	ATTGATCCGG	TAAATCGGTAT	GGCTCTGTGCA	GCAAAACCGAG
	5	CGGACAGCG	ATTATTATGG	CCAAATCGGA	TGCTTTGGGG	CGAG.ATTTT
	101	TGTTTTCCTG	GGAGTCGTATG	ACGCTCATGCG	TTCGGCATGCT	TTTGTGCTGTA
	151	TCATGATGTT	TTTGGTGTTT	TACACCAATT	TGTGCTCATG	TCGCATCGTG
55	201	CCGCGCTTCT	GGCGCGAATT	GAAGTCCTTT	CGGGAAGAAG	TAAAGAAATA
	251	ATCTCTGGCG	GCGATGCGCC	ATTCTTCGCT	TGTGGATGTA	AAAAATGGCG
	301	CCGAGATGCG	CAAGCTTAT	CTGGAAGATC	AGGTTTGTCA	GGGGAAACCC
	351	ATTACATCTG	GGGCGCTAT	CGGCTCGATT	AGGCTGCTAT	AGAGCTGCTAT
	401	GAACAAATGG	GCGTATATCT	CTGGACGCTAT	TTGCTTTGAT	TCATTCTGGT
	451	GGGCGCGGTT	GATAGACAAG	AACTCGTGGT	TGAAGATGGG	TATGCTGACC
60	501	TGTCGGATGT	TTCCGAGACA	TACAGCGGTT	TGTACCAAGG	ATTTC.AAGC

45	orf110.pep		LGIGASVIGTLLQNNQPQTDYLVKFGSFWA	30
	orf110.ng		MSKSRISPTLLSRPFAFFSSMRFAVALLSLGIGASVIGTVLQNNQPQTDYLVKFGPWT	60
50	orf110.pep		XIFGFLGLYDVYASAWFVVMFVLVSTSLCLIRNPFFWREMKSFREKVKESLAAMRH	90
	orf110.ng		RIFDFGLGLYDVYASAWFVVMFVLVSTSLCLIRNPFFWREMKSFREKVKESLAAMRH	120
	orf110.pep		SSLVDVKIAPEVAKRYLEVRGFGQGTINRSDGSVLAAKGGTMNKWGYIXAHVALIVCL	150
55	orf110.ng		SSLVDVKIAPEVAKRYLEVRGFGQGTIVSRDGSVLAAKGGTMNKWGYIXAHVALIVCL	180

orf110.pep	GGLIDSNLLKLGMLTGRIFRTIRRFMPRIKKPESEFGCVSLIXGQRQYFXRGVRVWF	210
	: : : :	
orf110ng	GRLINXNLLKLGMLAGSIFRNNRRVMPRIKSPESITGGVQSLITGQRQYFQRGKVRWF	240
orf110.pep	S 211	
orf110ng	S 241	

The complete length ORF110ng nucleotide sequence <SEQ ID 443> is predicted to encode a protein having amino acid sequence <SEQ ID 444>:

```

10      1  MSKSRISPTL  LSRPFAFFS  SMRFAVALLS  LLGIASVIGT  VLQONQPQTD
51     51  YLVKFGPFWT  RIFDFLLADH  VYASAVFVAT  MMFLVSVSTL  CLIRNVPVPE
101    101  REMKSFFKFK  KETSLAAMR  SSALDDVKAT  EVAKRYLEVR  GFQKTVTSRE
151    151  DGSVLIAARK  GKNSKWGITY  AHSLITVLCL  GRGLNXNLLL  KLGMLAGSIF
201    201  RNNRRVMPRT  STPSSWGWG  OATIKGROV  FORNGKRWMF  S*

```

15 Based on the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 53

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 445>:

20	1	ATGCGCTCTG	AAACAGCTGT	GGCGAACTTT	ATCCGGCTGT	TGATATTTGC
	51	CTCGGTTGTT	ATCTTGCTGA	ACGGCTGTTC	GGACAAACAT	GGCGAACACG
	101	TACCTCGTCA	GGGCGAARCG	ATGGGCAAGC	CCCTATACGGT	CAAAATACCTT
	151	TCAATTAATC	AGGCAACAAT	CCCTCCACGT	GGCGAAATAC	AAAAACGCAT
25	201	CGATGACGGG	CTTAAAGAA	CTAACACGCCA	GATGTGCCAC	TATCAGCCAGC
	251	ATCCGAAAT	CAGCCGGTTC	CAACCGACCA	CAGCGCGACC	GGCCCTCCGG
	301	ATTTCAAGCG	ATCTCGCACA	GGTTATCGCG	GAGCGCGTCC	GCTCGAACCG
	351	CTGCACACAC	GGGCGGGTGG	AGGTCAGTCC	GGCGCCCTTC	GTCAACCTTT
	401	GGGGTATGCG	TCGCGGATCA	CTGTTACAGC	GTGACACGAT	GGCGGACACA
30	451	ATCAACATCA	TCGCGGATCA	TGACGCAATC	GGCGGACGAT	GGCGGACACA
	501	AGGCAAGAAT	CCCGCTTCTT	TGACGCAAAC	CCACCCACAG	GCTCATTTGG
	551	ATTTATCTTC	GATTTGCCAA	GGCTTCGGCG	TGTATAAAGT	TGGCGGGGAA
	601	CTTGAAATAAT	ACGGCATCTA	AAATTAATCT	GTGGAATCTG	GGGCGGAGTT
	651	GCAGGCGCAA	GCGCAAAACG	CGCGGCGGCA	ACCGTGGCGG	ATCGTATATG
	701	AGCAGCCCAA	TATCGTCCAA	GGCGCACAAT	CGCAGATTAT	CGTCCGCTCG
35	751	AACAAACGTT	CGCTTGCAC	TTCCGGGAGT	TACCGTATTT	TCCAGCTGCA
	801	TAAGAACGCG	AAAGGCGTCT	CCCATATCAT	CAACCGGATC	AAACAACGAC
	851	CCATCAGCCA	CACCTCTGCC	TGCATCAGCG	TGTTGCGACA	CAGTGGGATG
	901	ACCGGCGGCA	CGCTTGTCAC	AGGATTATTC	GTATTTGGCG	AAACAACGAG
	951	CTTAAAGCGG	CAGCAGCGCG	AAACAACGCG	ATGTCGCGCG	ATGTCGCGCG
40	1001	ATAAAGCGGG	CTACGACACC	GCCATGTCTT	CCSAATTTGA	AAACCTGCTC
	1051	GGCTAA				

This corresponds to the amino acid sequence <SEO ID 446: ORF111>:

		MPSESTPLPMF	IRVLILFALGF	IFPNAQCEST	AQTVTLISGFT	MGTTYTVVKYL
45	51	NNNRDKLPSP	AEIQKRIDDA	LEKVNMCSPQ	YQPDSEISRE	NQHTAGKPLR
	101	ISSDFAHVTA	EAVRLNLRKH	DGLDVTGPE	VNLWGFGEKD	SVTREPSPEQ
	151	IKQAASTGTI	DKIILLQKGD	YASLSTHKPK	ALDLSLSIAK	GVFQKPVAGE
	201	LEKYIQNYVL	VEIGELGGLG	KGNARCEBPE	IGTEQPIVNO	GGNTQIVPVL
	251	NKRLSATSGD	YIRFVHKDNG	KRLSHIPN	NKPTSTSHLA	YSISVADAM
50	301	TADGLSTGLF	VIGETEALKL	AREKLAFLV	IVRDKGGYRT	AMSSEFEKLL
	351	R*				

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF111 shows 96.9% identity over a 351aa overlap with an ORF (ORF111a) from strain A of *N.*

meningitidis:

5	orf111a.pep	10	20	30	40	50	60
	orf111	10	20	30	40	50	60
10	orf111a.pep	70	80	90	100	110	120
	orf111	70	80	90	100	110	120
15	orf111a.pep	130	140	150	160	170	180
	orf111	130	140	150	160	170	180
20	orf111a.pep	190	200	210	220	230	240
	orf111	190	200	210	220	230	240
25	orf111a.pep	250	260	270	280	290	300
	orf111	250	260	270	280	290	300
30	orf111a.pep	310	320	330	340	350	
	orf111	310	320	330	340	350	
35	orf111a.pep	310	320	330	340	350	
	orf111	310	320	330	340	350	

The complete length ORF111a nucleotide sequence <SEQ ID 447> is:

40	1	ATGCGGCTCTG	AAACACGCCT	GCGAACTTT	ATCCGACACT	TGATATTTGC
	51	CCTGAGTTT	ATCTTCCTGA	ACGCTGTTC	GGACAAACG	GCGCAACCG
	101	TACCCCTGCA	AGGTGAAAG	ATGGGCACGA	CCTATACCGT	CRAATACCTT
	151	TCAAATATATC	GGGACNAACT	CNCNTCACT	GCGGAAATAC	AAANCGCAT
	201	CGATGACGGC	CTTAAAGAAG	TCAACCGGCA	GATGTCCACC	TATCAGCCCG
45	251	ACTCGGAAT	CAGCGGGTTC	AACCAACACA	CAGCGGGCAA	GCCCTCCCG
	301	ATTTCAGCG	ACTTCGCACA	CGTTACTGCC	GAGCGCGTCC	ACCTGACCCG
	351	CCTGACACAC	GGCGCGCTGG	ACGTAAACGT	CGCGCGCTGG	GTCAACCTT
	401	GCGGATTCGG	CCCGACAA	TCGCTACCG	TGCGACCGCT	GCGGACACAA
	451	ATCAACACAG	CAGCATCTTA	TACGGGCAAT	GACAAATCA	TTTTGAACAA
50	501	AGGCAAGAT	TACGCTTCT	TGACGAAAC	CCACCCCAAG	GCCTATTGG
	551	ATTATCTTC	GATTGCCAA	GGCTTCGGCG	TTGATNANGT	TGCGGGCGAA
	601	CTGGAAATAT	ACGGCATTC	AAATPATCTG	GTGAAATCG	CGCGNAGATG
	651	GCAAGCGCAA	GNCAAAAACG	CGCGCGGCGA	ACCTTGGCGC	ATCGGCATCG
	701	AACAGCCCAA	CATCGTCCAA	GCGCGCAATA	CGCAGATTAT	CGTCCCGCTG
55	751	AACAACCGTT	GNFTTGCACA	TTCGGCGAT	TACGCTATT	TCCACGTGCA
	801	TAAAGCGGC	AAACGCCCTC	CCCATATCAT	TAATCCGAC	AAACAACGAC
	851	CCATCAGCCA	CRACTCTGCC	TCCATCAGCG	TGNTCGCAGA	CAGTGCAGTG
	901	ACGCGCGACG	GCTTNTCCAC	AGGATTATTC	GTATTTGGCG	AAACCGAAGC
	951	CTTAAAGCTG	CGAGAGCGCG	AAAACCTGCG	TGTTTTCTTG	ATTGTGAGGG
60	1001	ATAAAGGCGG	CTACCGCACC	GCCATGTCTT	CGGAATTGGA	AAAACGTGCT
	1051	CGCTAA				

This encodes a protein having amino acid sequence <SEQ ID 448>:

1	MPSETRLPNF	IRTLIFALSF	IFLNACSEQT	AQTIVLQGET	MGTTYTVKYL
51	SNNRDLPSF	AEIQXRIIDA	LKEVNRQMST	YQPDSEISRF	NQHTAGKPLR

101 ISSDFAHVTA EAVHLNRLTH GALDVTVGPEL VNLWGFPGDK SVTIRESPFEQ
 151 IKQAASYTKL DKILKQSKD YASLSKSLAK GFGVDXVAGE
 201 LEKYGIQNYL VEIGELHKG KKNARGEPWR IGEQPNIVQ GGNTOIIVPL
 251 NNRSKATSGD YRIFHVDKSG KRLSHIINP NKRPIISHNLA SISVXADSAM
 5 301 TADGXSTGLF VLGETEALKL AEREKLAVFL IVRDKGGYKT AMSSEFEKLL
 351 R*

Homology with a predicted ORF from *N.gonorrhoeae*

ORF111 shows 96.6% identity over a 351aa overlap with a predicted ORF (ORF111.ng) from *N.*

10 *gonorrhoeae*:

	10	20	30	40	50	60
orfillng	MPSETRLPNLIRALIFALGFFILNACSECTAQTVTLQGETMGTTTYTVKYLNNRDKLPSP					
orfill1	MPSETRLPNFIIRVLIFALGFFILNACSECTAQTVTLQGETMGTTTYTVKYLNNRDKLPSP					
	10	20	30	40	50	60
	70	80	90	100	110	120
orfill1	AKIQKRIDDALKEVNRQMSYQTDSEISRFOHTAGKPLRISSEDFAHVTAEAVHLNRLTH					
20 orfill1	AEIQKRIDDALKEVNRQMSYQDPDSEISRFOHTAGKPLRISSEDFAHVTAEAVHLNRLTH					
	70	80	90	100	110	120
	130	140	150	160	170	180
orfillng	GALDVTVGPELVNLWGFPGDKSVTREPSEQIKQAASVTGIDKIIQQGKDYASLSKTHPK					
orfill1	GALDVTVGPELVNLWGFPGDKSVTREPSEQIKQAASVTGIDKIIQQGKDYASLSKTHPK					
	130	140	150	160	170	180
	190	200	210	220	230	240
orfillng	AYLDLSSIAKFGVDKVGAELEKYGIQNYLVEIGGELHKGKKNARGEPWRIGTQPNIVQ					
orfill1	AYLDLSSIAKFGVDKVGAELEKYGIQNYLVEIGGELHKGKKNARGEPWRIGTQPNIVQ					
	190	200	210	220	230	240
	250	260	270	280	290	300
orfillng	GGNTQIIVPLNNRSLATSGDYRIFHVDKNGKRLSHIINPNNKRPIISHNLASISVSDSAM					
orfill1	GGNTQIIVPLNNRSLATSGDYRIFHVDKNGKRLSHIINPNNKRPIISHNLASISVSDSAM					
	250	260	270	280	290	300
	310	320	330	340	350	
orfillng	TADGLSTGLFVLGETEALRLAEQKLAFLIVRDKDGYRTAMSEFAKLLRX					
orfill1	TADGLSTGLFVLGETEALRLAEQKLAFLIVRDKDGYRTAMSEFEKLLRX					
	310	320	330	340	350	

The complete length ORF111ng nucleotide sequence <SEQ ID 449> is:

1 ATGCGCGCTG AAACACGCGT GCCGAACCTT ATCCGCGCGT TGATATTTGG
 51 CCGTGGGTTT ATCTCTCTGA ACGCGCTGTC GGAaacaacC GCGCAaacgc
 101 TTACCTCTGCA AGGCGAaacg aTGGGTAAGA CTTATACGCT CAATACCTT
 50 151 TCAATAATAT CGGGCAAACT CCCTCCCTCT GCCAAATATC AAAAGCGCAT
 201 TGAATGATGCG CTTAAGAAGG TCAACCGGCA GATGTCACCC TACCAGACCG
 251 ATTCCGAAGT CAGCGCGTTC AACCAACACA CAGCCGGCAA GCGCCCTCCG
 301 ATTTCAGGCG ATTTGCGACA CGTTACCGCC GAAGCGGTCC GCCTGAACCG
 351 CCGTACTCAC GCGCGACTGG ACGTAACGCT CGGCCCTTTG GTCAACCTTT
 55 401 GGGGGTTCGG CCCGACAAA TCCGTTACCC GTGAACCGTC CGCGGAACAA
 451 ATCAACACAGG CGGCATCTTA TACGGCGATA GACAAATATC TTTTGCAACA
 501 AGGCRAAGAT TACGCTTCCT TGAGCAAAAC CCACCCAAA GCCTATTGGG
 551 ATTTATCTTC GATTGCCAAA GGCTTCGGCG TTGATAAGT TGCGGGCGAA
 601 CTGGAATAAT ACGGCATTTA AATATTATCT GTGAAATcg gggcgAGTT
 651 GCACGGCAAA GGCATAAATG CGCACGGCGA ACCGTGGGCG ATCGGTATAG
 701 AGCAACCCAA TATCATCCAA GgcgGCAata CGCAGATTAT cgtcccgctg
 751 aaCaacggtt cgtTGCCAC TTCGGCGAT TAcggtATT tccaagtcgA
 801 TAAAAacgg aaacgcctt ccaacaTAT CAATCCcAc acaAAcYac
 851 ccatCAGcga caactcgag tccatcagc tggctcAGA tggTcCAAtg
 65 901 ACGCGGACG GTTtatCCAC AGGATTATT GTTTTACGG AAACCGAAGC
 951 CTTAAGGCTG GCAGAACAGG AAAAATTCGC TGTTCCTTA ATTGTCGGG

1001 ATAAGGACGG CTACCGCACC GCCATGTCTT CCGAATTTCG CAAGCTGCTC
1051 CGCTAA

This encodes a protein having amino acid sequence <SEQ ID 450>:

1 MPSETRLPNL IRLALIFALGF IFLNACSEQT AQTVTLOGET MGTTYTVKYL
51 SNNRDKLPSP AKIQKRIDDA LKEVNRQMS YQTDSEISRF NQHTGAKPLR
101 ISSDFAHVTA RAVRLNRLTH GALDVTGVL VNLWGFPGDK SVTREPSPEQ
151 IKAQASYTGI DKII LQQQKD YASLSKTHPK AYLDLSSIAK GFGVDKVAKE
201 LEKYGIQNYL VEIGGELHGK GKNAHGEPR IGIEQPNIIQ GGNQIIVPL
251 NNRSLATSGD YRIFHVDKNG KRLSHIINPN NKRPISHNLA SISVVSADSAM
301 TADGLSTGLF VLGETEALRL AEQEKLAFLV IVRDKGYRT AMSSEFAKLL
351 R*

This protein shows homology with a hypothetical lipoprotein precursor from *H. influenzae*:

sp|P44550|YOJL_HAEIN HYPOTHETICAL LIPOPROTEIN HI0172 PRECURSOR >gil1074292|pir|4
hypothetical protein HI0172 - Haemophilus influenzae (strain Rd KW20)
15 >gil1573128 (U32702) hypothetical [Haemophilus influenzae] Length = 346
Score = 353 bits (896), Expect = 9e-97
Identities = 181/344 (52%), Positives = 247/344 (71%), Gaps = 4/344 (1%)
20 Query: 7 LPNLIRALIFALGFIFLNACSEQTACTVTLOGETMGTTYTVKYLNNRDKLPSPAKIQKR 66
+ LI +? + L AC +? + +L G+TMSTFY VKL + S K +
Sbjct: 1 MKKLSGIIAVAMALSAAQCQKET-KVISLSGKTMTSTTYHVKYLDGGSITATSE-KTHEE 58
25 Query: 67 IDDLAKVNRQMSYQTDSEISRFNQHT-AGKPLRISDFAHVTAEARLNLRLTHGALDV 125
I+ LK+VN +MSTY+ DSE+SRFNQ+T P+ IS+DFA V AEA+RLN++T GALDV
Sbjct: 59 IEAIIKDVNAKMSYKKDSELSRFNQHTVNTPIEISADFAKVLAEAIRLNKVTGALDV 118
30 Query: 126 TVGPLVNLWGFPGDKSVTREPSPEQIKQAAS YTGIDKII LQQQKDYASLSKTHKAYLDL 185
TVGP+VNLWGFGE+K ++P+PEQ+ + ++ GIDKI L K+A+LSK P+Y+DL
Sbjct: 119 TVGPVNLWGFGEPEKRPQPTPEQLAERQAWVGIDKITLDTNKEKATLSKALPQVYVDL 178
35 Query: 186 SSIKAGFGVDKVAKEKYGIQNYLVEIGGELHGKGNNAHGEPRIGIEQPNIIQGGNQ 245
SSIAKFGVD+VA +LE+ QNY+VEIGGE+ KGKN G+PW+I IE+P +
Sbjct: 179 SSIKAGFGVDQVAEKLEQLNAQNYMVEIGGEIRAKGNIEGKFWQIAIEKPTTTTGERAVE 238
246 IIVPLNNRSLATSGDYRI FHVDKNGKRLSHIINPNKRIFISHNLASISVVSADSAMTADGL 305
++ LNN +A+SGDYRI+ ++NGKR +H I+P FI H+LASI+V++ ++MTADGL
Sbjct: 239 AVIGLNMGMASGSDYRIY-FEENGKRFPAHEIDPKTGYPIQHHLASITVLAFTSMTADGL 297
40 Query: 306 STGLFVLGETEALRLAEQEKLAFLVIRVDKGYRTAMSSSEFAKL 349
STGLFVLGE +AL +AE+ LAV+LI+R +G+ T SS F KL
Sbjct: 298 STGLFVLGEDKALEVAEKNNLAVLYLIIRTNGFVTKSSSAFKKL 341

Based on this analysis, it is predicted that the proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 54

45 The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 451>:

1 ..CCGTGCGGCC GACAGGGCGA CGACGTGTAT CGGCGCGCAG CGTCCCGTCA
51 AAAATTGTGG CTGCGCTTCA TCGGCGCGCG GTGCGTACAA AATATACGGG
101 GCGGCGCGCG TCGGCGACGG TGGCGCAAGG CGGTGCAAAAT CGGCGCGGAG
151 GTGTTGTGAC GGCACAAATGA AGGCAGGCCA YTGCGCAATCG CGGTGATGGG
201 CGGCGAGGGCC GGCCAGCAGC CwTCAAGTCAA CGGCAAAAGCG GGTGCGGACG
251 GCACTGATTT GTATGGTTAT GgCGGGGgTG TTTATGCTGCG GTGCGATCAG
301 TCGCGCGCTA AACAAACGGG TjCGTATTTC GAGCGCTGCT TGCATACCA
351 ACGTTTCAAA CACCGCATCA ATATGAGAAA CCGCGCGGAA CgCTACAAA
401 CCAAGAGGTT GACGCGCTCT CTGGAAGGCG GCTACAACGC GTTGTGGGCG
451 GAAGGCGATTG TCGGAAAAGG CAATAATGTG CGGTTTTACC TACAACCGCA
501 GgCGCACTTT ACCTACTTGG CGGTAACCGG CGGTTTTACC GACAGCGAGG
551 GGACGCGCGT CGGACGCTGC GGCACGGGCT AGTGGCAAGG CCGCGCGCGG
601 ATTCGGGCAA AAACCGGTTT TGCTTTGCGT AACGGTGTCA ATCTTCAGCC
651 TTTTGCCTGT TTTAATGTTT TGCACAGGTC AAAATCTTTT GCGCTGGAAA
701 TGSACGGGGA AAAACAGAGC CTGGCAGGCA CAGGCGCACT CGAAGGCGGG

751 TTCGGTATTG AAGCCGGTTG GAAAGGCCAT ATGTCGCCA..

This corresponds to the amino acid sequence <SEQ ID 452; ORF35>:

```

1  ..PCRRGGDDVY AAHASRQKLW LRFITGRSHQ NIRGGAAADG WRKGVQIGGE
51  VEVFVONEGSX LAIGVMGGRA QHASVNGKG GAAGSDLYGY GGGVYAAWHQ
101 LRDKOTGAYL DGWLQYORFK HRINDENRAE RYTKGTWAS VEGGYNALVA
151 EGIVGKGNV RFYLPQPAQF TYLGVNGGFT DSEGTAVGLL GSGQWQSRAG
201 IRAKTRFALR NGVNLQPFAP FNVLHRSKSF GVEMDGEKQT LAGRTALEGR
251 FGIEAGWKGH MSA..

```

Computer analysis of this amino acid sequence gave the following results:

10 Homology with putative secreted VirG-homologue of *N. meningitidis* (accession number A32247)

ORF and virg-h protein show 51% aa identity in 261aa overlap:

```

Orf35 5 QGDDVYAAHASRQKLWLRFIGGRSHQNIIRGGAA-ADGWRKGVQIGGEVFRVONEGSXLA 63
+ D++ R+ LMLR I G S+Q ++G A +G+RKGVO+GGEVF QNE + L+I
15 virg-h 396 KNSDIFDRTLPRKGLMLRVIDGHSNQWVQKTAPEVGYRKGVLGGEVFTWQNESNQLSI 455
Orf35 64 GVMGGRAGQHASVNGKG--GAAGSDLYGYGGGVYAAWHQLRDKQTGAYLDGWLQYORFKH 121
G+MGGA Q ++ + ++ G+G GVYA WHQL+DKQTGAY D N+QYQRF+H
15 virg-h 456 GLMGGAQAEQRSTFHNPDTNLITGNVKGFGAGVYATWHQLDKQTGAYADSNWQYQRFH 515
20 Orf35 122 RINDENRAERYTKGTWASVEGGYNALVAEIGVKGKNNVRFYLPQPAQFTYLGVNGGFTD 181
RIN E+ ER+ +KG TAS+E GYNAL+AE KGN++R YLPQPAQ TYLGVNG F+D
15 virg-h 516 RINTEDGTERTPTSGKITASIEAGYNALLAEHFTKGNLSRVLYLPQPAQLTYLGVNGKFS 575
Orf35 182 SEGTAVGLLSSGQWQSRAGIRAKTRFALRNGVNLQPFAPFNVLHRSKSFVEMDGEKQTL 241
SE V LLGS Q Q+R G++AK +F+L + ++PFAP N L+ +K FGVEMDGE++ +
25 virg-h 576 SENAHVNLGSRQLQTRVGVQAKAQFSLYKNIATIEPFAAVNALYHNKPFVEMDGERRVI 635
Orf35 242 AGRTALEGRFGIEAGWKGHMS 262
+TA+E + G+ K H++
30 virg-h 636 NNKTAIESQLGVAVKIKSHLT 656

```

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF35 shows 96.9% identity over a 259aa overlap with an ORF (ORF35a) from strain A of *N.*

meningitidis:

```

35 orf35.pep 10 20 30
PCRRGGDDVYAAHASRQKLWLRFIGGRSHQNIIR
orf35a 10 20 30
ORLAIPEAEAVLYAQQAAYANTLFLGLRAADRGDDVYAADFSRQKLWLRFIGGRSHQNIIR
310 320 330 340 350 360
40 orf35.pep 40 50 60 70 80 90
GAAADGWRKGVQIGGEVFRVONEGSXLAIGVMGGRAGQHASVNGKGGAAGSDLYGYGGGV
orf35a 40 50 60 70 80 90
GAAADGRRKGVQIGGEVFRVONEGSRLAIGVMGGRAGQHASVNGKGGAAGSYLHGYGGGV
370 380 390 400 410 420
50 orf35.pep 100 110 120 130 140 150
YAAWHQLRDKQTGAYLDGWLQYORFKHRINDENRAERYTKGTWASVEGGYNALVAEGIV
orf35a 100 110 120 130 140 150
YAAWHQLRDKQTGAYLDGWLQYORFKHRINDENRAERYTKGTWASVEGGYNALVAEGV
430 440 450 460 470 480
55 orf35.pep 160 170 180 190 200 210
GKGNVRFYLPQPAQFTYLGVNGGFTDSEGTAVGLLSSGQWQSRAGIRAKTRFALRNGVN
orf35a 160 170 180 190 200 210
GKGNVRFYLPQPAQFTYLGVNGGFTDSEGTAVGLLSSGQWQSRAGIRAKTRFALRNGVN
490 500 510 520 530 540
60 orf35.pep 220 230 240 250 260
LQPFAPFNVLHRSKSFVEMDGEKQTLAGRTALEGRFGIEAGWKGHMSA
|||||

```

orf35a LQPFPAFFNVHLHRSKSGFVEMDGEKQTLGRTALEGRFGIEAGWKGHMSARIGYGKRTDGD
550 560 570 580 590 600

5 orf35a KRAALSLKWLFX
610 620

The complete length ORF35a nucleotide sequence <SEQ ID 453> is:

	1	ATGTTCCAGAT	CTCAGCTTGG	TTCAAAATCT	GTTTCTACCA	AAATCGCGCA
	51	CGATGCGGAG	TTTTCATTTT	CAGACAGGCC	GAACCCGGCG	ATCTCCCATT
10	101	ATTTTTCGAC	CGGTAAAAAC	GATCAAATAT	CTACCGTATA	TGGGTATGAC
	151	GAATCAATA	TCCAAGGATA	ATAACTAAT	ACGGCGATAC	TGCGCGTCGA
	201	TAAATTTGCC	GTGTTTAAAG	AATATATAAT	AGATACTTAC	GGGGATAATT
	251	TAAAGGATGC	GGTTAAGGAT	CATTATCAGG	ATTATTACAA	ACAAGACCCC
	301	GAAAGCTTGG	AAGAAATATA	AAANCGAAT	GGAGAGCGCT	ATATAGAACT
	351	GCTTGTACCA	AAAAATTAGT	TACTCAAACT	GA AAAACCCC	GATTTAATTAT
15	401	ATAAATTGTT	AGAAAGATT	CTACTCATC	CTCATGTAGT	TACATCCACAG
	451	ACTAGTCTGT	ACAACATCTT	TATCAAAAA	TACACGTCTA	AAATCGAAAA
	501	CAATATCCAC	GTGCGCGGAC	AGGTTGTGGA	ATTGACACAG	ATGACGCTGA
	551	ARGATTCCTT	TGGGAGCCAG	CGCGCCATCT	CCGACATCCA	TATGCTGGAA
	601	ATCTTCGATA	TGCGCCGATC	CGCGCTGCA	ACGAAAGATG	AAAAATCGAC
20	651	CGTCCATATA	CGGTATAGG	CGTGTGGAAG	TCTTTGTCTC	GGCTACAGAG
	701	TGCGGAGATG	CGTCAAAAGC	CGCTCAAACT	AGTGGTGGAT	AGTGGTGGAT
	751	CAATCCGCGC	TGGTTTGGGA	ACGCGCGCGC	GAATACTGTA	AAAGCTGTGA
	801	CGCGCGCAAA	CTGATCTGAG	CGAAGAAAGC	AGACTTAAT	TGCTTTGGCT
25	851	TAAACAAAA	TTACCGGCGC	GGACTGTAGC	ATTATTATCT	CAAGCAATCT
	901	GAAAGCGGAT	TTGCTTTGGG	GCTGCAGCGT	TTGGCTATGC	CGAGGCGCGA
	951	AGCGGTTTTA	TGCTCCCAAC	AGGCTTATGC	GGCAATATCT	TTGTTCCGGC
	1001	TGCGTGGCGC	CGACAGCGGAG	GACGACGTGT	ATGCGCTAGC	TCGCTCCCGT
	1051	CAAAAATTGT	GGCTTGCGCT	CTATCGGCGC	CGGTGCGCAT	AAAAATATCG
30	1101	GGGCGGCGCG	GCTCGGAGAC	GGGCGGCGAA	AGGCGTGCAT	ATCGGCGAGC
	1151	AGGTTGTGTT	ACGGCAAAAT	GAAAGCCAGC	GGCTCGCATC	CGCGCTGATG
	1201	CGGCGCAGGG	TGCGGCACGA	CGCATCAGTC	ATGCGCAAAG	CGGCTGGCGT
	1251	AGGCAGTATT	TTGCTATGTT	ATGCGGGGGG	TGTTTATGTCT	CGCTGGCATC
	1301	AGTTTGCGGTA	TAAAAACAA	GTTGCGTATT	TGGAGCGCTG	GTTCACATAT
	1351	CACGTTTCTA	AACACGCGAT	CAATGATGAT	AACCGTGGAC	AACGCTACAA
35	1401	ACCAACAAAT	TGGACGGGTT	CTGTGGAAG	CGGCTAGCG	GCSTTTGGTG
	1451	CGGAGGAGCT	TGTCGAAAA	CGGCATTAAT	TGCGGTTTTC	CTTGCACACG
	1501	CGGAGCGGAT	TTGCGGATCT	GTGCGGATCT	TCAGTGTGTA	ACGCGCGCGC
	1551	GGGAGTGGCG	GTGCGGATCT	GTGCGGACGC	TCAGTGTGTA	ACGCGCGCGC
	1601	GCATTGGGCG	AAAAACCCGT	TTTGTCTTGC	GTAAACGTTT	CAATCTTCAG
40	1651	CGTTTGGGCG	CTTTTAATGT	TTTGGACAGG	TCACAAAGCT	TGCGGCTGCT
	1701	AATGACGACG	GA AAAACAGA	CGCTGCGACG	CAGGACGGCG	CTCGAAGGCG
	1751	GGTTCCGCGAT	TGAAGCCGGT	TGGA AAAGCC	ATATGTCGCG	ACGCAATCCG
	1801	TACGCGAAAA	GTGCGGACGG	CGACAAAGAA	GCGCATATTG	CGCTCAAAAT
	1851	GCTGTTTGA				

45 This encodes a protein having amino acid sequence <SEQ ID 454>:

	1	MFRQAQSGNT	RSTIGDGDAD	PSFVKDKPKPG	TSHYPSVSKT	DONSSEYGYD
	51	ENIQOGKNYN	SGILAVDMND	VVVKYITDYT	GDNLKDSVKK	QLQDLYKTRP
	101	EAWTEENKRT	EAYEYBQLGP	FLKQKQKNP	DLINKLVEDS	VLTPHSNTSQ
50	151	TSLNINFNKK	LHVIKESNKN	VAGVLELTKE	MTLKDSLWEP	RRHSDIHMLE
	201	TSDNARILRN	TKDEKTLVHK	YAQGGADFNS	GYVDRESDKT	ALTFFEEKVSG
	251	QSGVULERP	ENLKTLDGRK	LIAAEKADSL	SPFAVNYRQ	GLYLELHKQC
	301	EGGFCFLQVR	LAIPEAEAVL	YAQQAQYVGT	LFGLRAARDN	DDVYAADAPR
	351	QKMLRWLGF	RSHGNTRGGA	AADGRKAGNT	IFGSEVFRQN	EGSRJALPSM
55	401	QGRACQWQK	LHGLRQKQK	LHGLRQKQK	ALGLDGLWQK	QALDGLWQK
	451	QRFKHLEND	NRAEKVYKES	WTASVEGQYN	ALVAGVAKT	GRAGTRVTRK
	501	QAQPTYLGVN	GGFTDSECTA	VGLSGSGQYN	SRAGTRVTRK	PAIRNGNVLQ
	551	PPAFANVLHR	SKSFGVEMDG	EKQTLAGRTA	LEGFRGFIAG	WKGHMSARTG
	601	YGRKTDGDKK	AALSLKDWLF*			

Homology with a predicted ORF from *N.gonorrhoeae*

60 ORF35 shows 51.7% identity over a 261aa overlap with a predicted ORF (ORF35ngh) from *N.*

gonorrhoeae:

```

orf35.pep                                PCRROGDDVYAAHASRQKLWLRFIGGRSHQNIHG 34
                                         ::|::      |::||| |::|::|
orf35.ngc    FTKVOERDDIAIYAQAQAANTLFALRNDKNSDIFDRTLPRKGLWLRVIDGHSNQWVQG 370

```


451 AAAGGCGTGT CCCAAATAT CGCCTCAAC CTGACCGACA ACCGAGCAGC
501 CGGACACAGG CTGCGCAGCC GTTCCACAA TSCCGGTACT ATGCTCAACG
551 AAGGAGTAGG CGACGCATTC AACCAGGCGCA CCGATACAG CCCGAGCGTG
601 GACAGATCGG GCAATGCGG CSAAGCTTTC AACGCGACTG CAGATATCGT
5 651 TAAAAACATC ATCGGCGCTG CAGGAGAAAT TGT

This corresponds to the amino acid sequence <SEQ ID 460; ORF46-1>:

1 ..AVCLPMHAHA SXLANDSFIR OVLDROHFEP DGKHYLFGRS GELAEORSHI
51 GLGKIQSHOL GNLMIQQAII KGNIGYIVRF SDHGHEVHSP FUNHASHISDS
101 DEAGSPVDGF SLRYIHWGDI EHHFPADYDG PQGGGYPAKP GARDIYSYDI
151 EGVQNIIRLN LTDNRSTGOR LADRFHNAGS MLTQGVGDGF KRATRYSPEL
201 DRSGNAEEAF NGTADIVKNI IGAAGEI

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N. gonorrhoeae*

ORF46 shows 98.2% identity over a 111aa overlap with a predicted ORF (ORF46ng) from *N.*

15 *gonorrhoeae*:

orf46.pep	AEYVQFSIDLFSVKGSGGGIPKAKPVFDAKPRWEVDRLKNLKTTR	45
orf46ng	PKTGVFPDGGKFPNFEXHVKYDTKLDIQLSGGGIPKAKPVFDAKPRWEVDRLKNLKTTR	217
20 orf46.pep	EQVEINVQETRRRSQSSQFKAHQAQREWENKTGLDNHFIIGGDKNGKGTVTGGHSLTRGDV	105
orf46ng	EQVEINVQETRRRSQSSQFKAHQAQREWENKTGLDNHFIIGGDKNGKGAVTGGHSLTRGDV	277
25 orf46.pep	RVIQQTAPDKHGXLSSDSGN 126	
orf46ng	RVIQQTAPDKHGVLSSDSGN 298	

A partial ORF46ng nucleotide sequence <SEQ ID 461> is predicted to encode a protein having partial amino acid sequence <SEQ ID 462>:

30 1 ..RLKHCCHAR LGSAPHRQD GAHQRFGRYQ ATQRLCRSSH PRIGSPKPCQ
51 RTRHRSRQOY LYGSHPHQD WSPCPGKIQLG RHGTSRCRAV ADXRDRICER
101 EIRQRQXCR CRLGKIPSL S IPKYPLKLEQ RYKENITSS TVPPNSGNKV
151 KLDQRHPKT GVPFDGCGFF NFEKHVRYDT KLDIQLSSGG GIPKAKPVFD
201 AKPRWEVDRL LNKLTTRQV EKNVQETRRK SQSSQFKAHA QREWENKTGL
251 DFNHFIIIGDI NKGAVTGGH SLTRGDVRI QQTSAPOKGG VLSSDSGN*

35 Further work revealed the complete gonococcal DNA sequence <SEQ ID 463>:

1 TTGGGCATTTC CCCGCAAAAT ATCCCTTATT CTGTCCATAC TGGCAGTGTG
51 CCTCGCGNTG CATGCACACG CCTCAGATTG GGcaAACGAT CCCTTTATCC
101 GgCaggttct CGaacCTCAG CMTTTCGaac ccgacggGaa ATACCaCCTA
151 TTcggCaGCa GGGGGGAGCT TgcnagcGcc aacggccAta tcggattggG
40 201 aaacaTAcAa Agccatcagt tgggccacat gatgatcaaa caggcgccgC
251 ttgaaggaaA TAtcgGctac attgtccgct tttccgataca cgggcacaaa
301 ttccattcgc ctttcGAcAa ccaTGCTTCA CATTCCGATT CTGACGAAGC
351 CGGTAGTCCC GTTGACGGAT TCAGCCTTTA CCGCATCCAT TGGACGGAT
45 401 ACGAACACCA TCCCGCGGAC GGCTATGAC GGCACAGGG CGCGCGCTAT
451 CCCGCTCCCA AAGGCGCGAG GGATATATAC AGCTACGACA TAAAGGCGT
501 TGCCCAAAAT ATCCGCCTCA ACCTGACGCA CAACCGCAGC ACCGCAACAC
551 GGCTTGCGGA CGGTTTCCAC AATGCGCGCG CTATGCTGAC GCAAGGAGTA
601 GGGCAGCGAT TCAAAAGCGC CACCCGATAC AGCCCCGAGC TGGACAGATC
651 GGGCATTGcc gccGAAGCTC TCAACGCGAC TGCAGATATC GTCAAAAACA
50 701 TCATCGGCGC GGCAGGAGAA ATTGTCGGCG CAGGCGATGC CGTGCGagGT
751 ATAAGCGAAG GCTCAAAACAT TGCTGTGATG CAGGCTTGG GTCTGCTTTT
801 CACCGAAAC AAGATGGCGC GCATACAGCA TTTCGAGAT ATGGCGCAGC
851 TCAAGACTA TCGCGCAGCA GCGATCGCG ATTGGCGCAT CCRAACCCG
901 AATGCGGAC AAGGCATAGA AGCCGTACAG AATATCTTTA TGGACGACAT
951 CCCATCAAA GGGATTGGAG CTGTCCGGGG AAAATACGGT TTGGCGGCA
55 1001 TCAACGACCA TCCTGTCAAG CCGTCGAGA TGGCGCGGAT CGCATTCGCG
1051 AAAGGGAAAT CCGCGCTCAG CGACAATTTT GCGGATGCG CATACGCCAA
1101 ATACCGGTCC CTTTACCATT CCGCAAAAT CGGTCAAAAC TTGAGACGAC

1151 GTTACGGCAA AGAAAACATC ACCTCCTCAA CGCTGCCGCC GTCAAACGGC
 1201 AAAAATGTCA AACTGCGAC CCGTCCGCCA CGCAAGACAG GCCTACCGTT
 1251 TCACGGTAAA GGGTTTCCGA ATTTTGTAGAA GCACGTGAAA TATGATACGA
 1301 AGCTCGATAT TCAAGAATTA TCGGGGGGCG GTATACCTTA GGCTAAGCCT
 1351 GTGTTTGTATG CGAAACCGAG ATGGGAGGTT GATAGGAAGC TTAATAAAT
 1401 GACAACTCGT GAGCAGGTGG AGAAAAATGT TCAGGAAAGC AGAAGAAGGA
 1451 GTCAGAGTAG TCAGTTTAAA GCCCATCGCG AACGGAATGT GGAATAAATA
 1501 ACAGGGTTAG ATTTTAAATCA TTTTATAGGT GGTGATATCA ATAGAAAAAG
 1551 CACAGTAACA GGAGGGCATA GTCTAACCCG TGGTGATGTA CGGTGATAC
 1601 AACAAACCTC GGCACCTGAT AACATGCGG TTTATCAAGC GACAGTGGAA
 1651 ATTAATAAAGC CTGATGGAG TGGGAGGTG AAAACGAAA AAGGTGGGAA
 1701 AGTGATGACC AAGCAGACCA TGTTCGCAAA AGATTGGGAT GAGGCTAGAA
 1751 TTAGGGCTGA AGTTACTTCG GCTTGGGAAA GTAGAATAAT GCTTAAGGAT
 1801 AATAAATGGC AGGTACAAAG TAAATCGGTT ATTAATAATAG AAGGATTTC
 1851 CGAACCTAAT AGAACAGCAT ATCCATTTA TGAATAG

This corresponds to the amino acid sequence <SEQ ID 464; ORF46ng-1>:

1 LGISRKISLI LSILAVCLPM HAHASDLAND PFIRQVLDQR HFEPDQGYHL
 51 FGSRGELAXR NGHILGNIQ SHQLGHLMIQ QAAVEGNIGY IVRFSDHGKH
 101 FHSFPMNAS YSDHSGEASP VDGFSLYRIH WDGYEHPAD GYDGPQGGY
 151 PAPKGARDIY SYDQKVAQN IRLNLTDNRS TGQRLADRFH NAGAMLTQDNR
 201 GDGFKRATRY SPELDRSNGA AEAENGADIV VKNIIGAAE IVGAGDAVQG
 251 ISEGSNIAMV HGLGLLSTEN KMARINDLAD MAQLKDYAAA AIRDWAVQNP
 301 NAAQGETEAS NIFMAAIPK IGAVRGKYG LGGITAPVVK RSMGAJALP
 351 KGKSAVSDNF ADAAYAKYPS PYHSRNIRSN LEQRVYKEMI TSSTVPPSPG
 401 KNIVLADQRH PKTGVPFDGK GPFNFEKHVK YDTKLDIQEL SGGGIPKAKP
 451 VFDAPRWEV DRKLNKLTR EQVEKNVQET RRRSSSQFKA AHAQREWENK
 501 TGLDFNHFIG GDINKKGTVT GGHSLTRGDV RVIQQTSA PD KHGVYQATVE
 551 IKKPDGSEWE KTKKGGKVM THTMFPKDW EARIAREVTS AWESRIMLKD
 601 NKWQGTSGSK IKIEGFTEPN RTAYPIYE*

ORF46ng-1 and ORF46-1 show 94.7% identity in 227 aa overlap:

		10	20	30	40		
orf46-1.pep		AVCLPMHAHASXLANDSFIRQVLDQRHFEPDQGYHLFSGRGEALAE					
orf46ng-1	LGISRKISLILSILAVCLPMHAHASDLANDPFIRQVLDQRHFEPDQGYHLFSGRGEALAE	10	20	30	40	50	60
orf46-1.pep	QSHIGLGIKTSQSHQLGNLMIQAAIKGNIGYIVRFSDHGHEVHS PFDNHASHSDSDEAGSP	50	60	70	80	90	100
	::: : : : : : : : : :						
orf46ng-1	NGHILGNIQSHQLGHLMIQAAVEGNIGYIVRFSDHGHEVHS PFDNHASHSDSDEAGSP	70	80	90	100	110	120
orf46-1.pep	VDFGFSLYRIHWDGYEHHPADGYDGPQGGGYPA PKGARDIYSYDIKGVAQNIRLNLTDNRS	110	120	130	140	150	160
orf46ng-1	VDFGFSLYRIHWDGYEHHPADGYDGPQGGGYPA PKGARDIYSYDIKGVAQNIRLNLTDNRS	130	140	150	160	170	180
orf46-1.pep	TGQRLADRFHNAGSMLTQGVGDGFKRATRYSP ELDRSNGAAEAENGADIVKNIIGAAE	170	180	190	200	210	220
orf46ng-1	TGQRLADRFHNAGAMLTQGVGDGFKRATRYSP ELDRSNGAAEAENGADIVKNIIGAAE	190	200	210	220	230	240
orf46-1.pep	I						
orf46ng-1	IVGAGDAVQGISSEGSNIAMVHGLGLLSTENKMARINDLADMAQLKDYAAAAAIRDWAVQNP	250	260	270	280	290	300

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF46ng-1 shows 87.4% identity over a 486aa overlap with an ORF (ORF46a) from strain A of *N. meningitidis*:

10 20 30 40 50 60

	orf46a.pep	LGISRKISLILSLAVCLPMHAHASDLANDSFIRQVLDROHFE PDGKYHFGSRGELAE R
	orf46ng-1	LGISRKISLILSLAVCLPMHAHASDLANDSFIRQVLDROHFE PDGKYHFGSRGELAXR
5		10 20 30 40 50 60
	orf46a.pep	SGHIGLGNIQSHQLGNLFICQAAIKNGIYIVRFSDHGHEVHSPFNHSHSDSDSDEAGSP
	orf46ng-1	NGHIGLGNIQSHQLGHLIMICQAAVEGNIGYIVRFSDHGKHFS PFNHNHSDSDSDEAGSP
10		70 80 90 100 110 120
	orf46a.pep	VDGFSLYRIHWGDYEHHPADGYDGPQGGYFAPKGARDIYSYDIKGV AQNIRNLNTDNRS
	orf46ng-1	VDGFSLYRIHWGDYEHHPADGYDGPQGGYFAPKGARDIYSYDIKGV AQNIRNLNTDNRS
15		130 140 150 160 170 180
	orf46a.pep	TGQRLVDRFHTGSMLTQGVGDGFKRATRYSPELDORSNAAEAFNGTADIVKNIIGAAGE
	orf46ng-1	TGQRLADRFHNAGAMLTQGVGDGFKRATRYSPELDORSNAAEAFNGTADIVKNIIGAAGE
20		190 200 210 220 230 240
	orf46a.pep	IVGAGADVQGISSEGSNIAMHGLGLLSTENKMARINDLADMAQLKDYAAAIRDWAVQNP
	orf46ng-1	IVGAGADVQGISSEGSNIAMHGLGLLSTENKMARINDLADMAQLKDYAAAIRDWAVQNP
25		250 260 270 280 290 300
	orf46a.pep	NAAQGIEAVSNIFTAVIPVKIGAVRGYKGLGGITAHFPVKRSQMGBIALPKGSASVSNDF
	orf46ng-1	NAAQGIEAVSNIFMAAIPKIGIGAVRGYKGLGGITAHFPVKRSQMGAIALPKGSASVSNDF
30		310 320 330 340 350 360
	orf46a.pep	ADAAAYAKYPSFYHSRNRISNLEQRYKENITSSVTPPSNGKNVKLANKRHPKTKVPFDCK
	orf46ng-1	ADAAAYAKYPSFYHSRNRISNLEQRYKENITSSVTPPSNGKNVKLADQRHPKTKVPFDCK
35		370 380 390 400 410 420
	orf46a.pep	GFNPFKDVKYDTRINTAVFQVN----PIDEPVEN--PKGSVGSARHSWSITARIQYAKLP
	orf46ng-1	GFNPFKDVKYDTKLD--IQELSGGGIPKAKPVFDAKPRWEVDRLKN-KLTTREQVEKRV
40		430 440 450 460 470
	orf46a.pep	RQGRIRYIPFKNYSAPSAPLPGKPNNGYLDKFGNEWTKGPSRTKGQFEDWDLSKGTGREQ
	orf46ng-1	QETRRRSQSSQFKAHQAQREWKNTGLDFNHFIIGDINKKGTVTGGHSLTRGDVRIQQTS
45		480 490 500 510 520 530
	orf46a.pep	CGGTAGTCCC GTTGACGGAT TCAGCCTTTA CGCATCCAT TGGGACGGAT
	orf46ng-1	CGGTAGTCCC GTTGACGGAT TCAGCCTTTA CGCATCCAT TGGGACGGAT
50		540 550 560 570 580 590
	orf46a.pep	CGGTAGTCCC GTTGACGGAT TCAGCCTTTA CGCATCCAT TGGGACGGAT
	orf46ng-1	CGGTAGTCCC GTTGACGGAT TCAGCCTTTA CGCATCCAT TGGGACGGAT
55		600 610 620 630 640 650
	orf46a.pep	CGGTAGTCCC GTTGACGGAT TCAGCCTTTA CGCATCCAT TGGGACGGAT
	orf46ng-1	CGGTAGTCCC GTTGACGGAT TCAGCCTTTA CGCATCCAT TGGGACGGAT
60		660 670 680 690 700 710
	orf46a.pep	CGGTAGTCCC GTTGACGGAT TCAGCCTTTA CGCATCCAT TGGGACGGAT
	orf46ng-1	CGGTAGTCCC GTTGACGGAT TCAGCCTTTA CGCATCCAT TGGGACGGAT
65		720 730 740 750 760 770
	orf46a.pep	CGGTAGTCCC GTTGACGGAT TCAGCCTTTA CGCATCCAT TGGGACGGAT
	orf46ng-1	CGGTAGTCCC GTTGACGGAT TCAGCCTTTA CGCATCCAT TGGGACGGAT
70		780 790 800 810 820 830
	orf46a.pep	CGGTAGTCCC GTTGACGGAT TCAGCCTTTA CGCATCCAT TGGGACGGAT
	orf46ng-1	CGGTAGTCCC GTTGACGGAT TCAGCCTTTA CGCATCCAT TGGGACGGAT

The complete length ORF46a DNA sequence <SEQ ID 465> is:

55	1	TTGGGCATT	CCGCAAAAT	ATCCCTTATT	CTGTCATAC	TGGCAGTGTG
	51	CTGCGCATG	CATGCACAG	CCTCAGATT	GGCAACAGAT	TCITTTATCC
	101	GGCAGGTCT	CGACCGTCAG	CATTTCGAAC	CCGACGGGAA	ATACCACTTA
	151	TTCCGCAGCA	GGGGGAACT	TGCGGAGCGC	AGCGGTGATA	TCGATTGGG
	201	AAACATACAA	AGCCATCAGT	TGGGCAACCT	GTTCCATCAG	CAGGCGGCCA
	251	TTAAAGGAAA	TATCGGCTAC	ATTGTCCGCT	TTTCGATGTA	CGGCGACGAA
60	301	GTCCATTTCC	CCTTCGACAA	CCATGCTTCA	CATTCCGATT	CTGATGAAGC
	351	CGGTAGTCCC	GTGACGGAT	TCAGCCTTTA	CGCATCCAT	TGGGACGGAT
	401	ACGAAACCA	TCCGCGCGAC	GGCATATGAC	GGCCACAGGG	CGGCGGCTAT
	451	CCGCTGCCA	AAGGCGCGAG	GGATATATAC	AGCTACGACA	TAAAAGCGGT
	501	TGCCCAAAAT	ATCCGCTTCA	ACCTGACCGA	CAACCGCGAC	ACCGGACAAC
65	551	GGCTTGTGCA	CGGTTTCCAC	AATACCGGTA	GTATGCTGAC	GCAAGGAGTA
	601	GGGACGGAT	TCAAAGCGGC	CACCGGATAC	AGCCCGGAGC	TGGACAGATC
	651	GGGCAATGCC	GGCGAAGCTT	TCAACGCGAC	TGCAGATATC	GTCAAAACAA
	701	TCTCCGCGC	GGCAGGAGAA	ATTGTCCGCG	CAGGCGATGC	CGGCGCGGTT
	751	ATAAGCGAAG	GCTCAACAT	TGCTGTTATG	CAGGCGTTGG	GTCTGCTTTC
70	801	CACCGAAGAC	ANGATGGCGC	GCATCAACGA	TTTGCAGAT	ATGGCGCAAC

851	TCAAAGACTA	TGCCGCAGCA	GCCATCCGG	ATTGGGCGAT	CCAAACCCG
901	AATGCCGAC	AAGGCATAGA	CTCTCTCAGC	AAATCTTTA	CGCGAGTCAT
951	CCCCGACAA	GGATTGGAG	CTCTCGGGG	AAAATACGGT	TTGGGGCGCA
1001	TGACGGCACA	TCTGTGCAAG	CGGTCCGAGA	TGGGGGAGAT	CGCATTGCGG
1051	AAAGGGAAAT	CCGCGCTCAG	CGACAATTTT	CGCGATCGGG	CATACGCCAA
1101	ATACCCGTC	CCTTACCATT	CCGAAATAT	CGTTTCAAC	TTGGAGCAGC
1151	GTTCACGGCA	AGAAACATC	ACCTCCTCAA	CGGTCCGGC	GTCAAACGGA
1201	AAGAAATGTA	AACTGGCAAA	CAACGCCAC	CCGAAGCCA	AAGTGCCGTT
1251	TGACGGTAAA	GGGTTTCCGA	ATTTTGAAAT	AGACGTAAA	TACGATACGA
1301	GAATTAATAC	CGCTGTACCA	CAAGTGAATC	CTATAGATGA	ACCGTCTTTT
1351	AATCCTAAAG	GTTCTGTCCG	ATGGGTCAT	TCTTGGTCTA	TAACTGCCAG
1401	AATTCAATAC	GCRAAATTAC	CAAGGCAAGG	TAGAATCAGA	TATATCCAC
1451	CTAAAAATTA	CTCTCCTTCA	GCACCGCTAC	CAAAAGGACC	TAAATATGGA
1501	TATTTGGATA	AATTTGGTAA	TGAATGGACT	AAAGTCCAT	CAAGAACTAA
1551	AGGTCAAGAA	TTTGAATGG	ATGTTCAATT	GTCTAAAAAC	GGAAGAGAGC
1601	AACTTGGATG	GGCTAGTAGG	GATGTTAAGC	ATTAAATAT	ATCAATTGAT
1651	GGAAGATTAT	CACACAAATG	A		

This corresponds to the amino acid sequence <SEQ ID 466>:

1	LGISPKISLI	LSILAVCLPM	HAHASDLAND	SFRQVLDRO	HFEPDGKYLH
51	FGSBEGLAER	SCHILGNLIG	SHQLGNLFIO	QAARICNIGY	IIVFSDHIGHE
101	VHSFPDNHAS	HSDSDEAGSP	VDGFSLYRIH	WDGYEHHPAD	GVDGQGGGY
151	PAPKARDIY	SYDIKQVAQN	IRLNLTDNRS	TGQRLVDRPH	NTGSMILTQGV
201	GDGFKRATRY	SPELDRSGNA	AEAFNGTADI	VKNIIAAGE	IVGAGDAVQG
251	ISEGNSIAVM	HGLGLLSTEN	KMARINDLAD	MAQLKDVAFA	AIRDWAVQNP
301	NAAQGIEAVS	NIFTAVIPVK	GIGAVRGKYG	LGGITAHFVK	RSQMGEIALP
351	KGKSAVSDNF	ADAAAYAKPS	PYHSNRNRSN	LEQRYGKENI	TSTVPPSPNG
401	KNIVLANKRH	PKTKVFPDGL	GFPNFEKDVK	YDRINTAVP	QVNPIDEPVF
451	NPKGSVGSAA	SWSITARIQY	AKLPRQGRIR	YIPPKNYSPS	APLPKGFNNG
501	YLDKFGNEWT	KGFSRTKGQE	FENDVQLSKT	GREQLGWASR	DGKHLNISID
551	GKITHK*				

Based on this analysis, including the presence of a RGD sequence in the gonococcal protein, typical of adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

35 Example 56

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 467>:

1	ATGAATATTC	ACACCCCTGCT	CTCCAAACAA	TGGACGCTGC	CGCCATTCTCT
51	GCCGAACGG	CTGCTGCTGT	CCTCTGCTGAT	ACTGCTTGCC	CCCAATCGGG
101	TGTTTGGGT	TTTGGCACTG	CTGACCGCCA	CGCGCCGCC	GATTGTCAAT
151	TTGGACTATC	TTCGCCCGCG	GCTGCTGATC	CGCCTGCCTT	GGCGTTTCTG
201	CAAAATTGCC	GGCGTATTGG	CGTTTGGGCT	GGCGGTTTGT	TTTGACGGGC
251	TGATGATGGT	GATCCAACTC	TTCCTTTTAA	TGGATCTCAT	CGGCGCCATC
301	AACTCGTCC	CCTTCATCCT	GACCGCCCC	GCCCCCTATC	AGATAATGAC
351	CGGGCTG...				

45 This corresponds to the amino acid sequence <SEQ ID 468; ORF48>:

1	MNIHTLLSKQ	WTLPPFLPKR	LLLSLLILLA	PNAVEWVIAL	LTATARPPIV
51	LDYLPALLI	ALPWFVKIA	GVLPFWLAVL	FDGLMMVIQL	PFPMDLIGAI
101	NLVPFILTAP	APYQINTGL...			

Further work revealed the complete nucleotide sequence <SEQ ID 469>:

1	ATGAATATTC	ACACCCCTGCT	CTCCAAACAA	TGGACGCTGC	CGCCATTCTCT
51	GCCGAACGG	CTGCTGCTGT	CCTCTGCTGAT	ACTGCTTGCC	CCCAATCGGG
101	TGTTTGGGT	TTTGGCACTG	CTGACCGCCA	CGCGCCGCC	GATTGTCAAT
151	TTGGACTATC	TTCGCCCGCG	GCTGCTGATC	CGCCTGCCTT	GGCGTTTCTG
201	CAAAATTGCC	GGCGTATTGG	CGTTTGGGCT	GGCGGTTTGT	TTTGACGGGC
251	TGATGATGGT	GATCCAACTC	TTCCTTTTAA	TGGATCTCAT	CGGCGCCATC
301	AACTCGTCC	CCTTCATCCT	GACCGCCCC	GCCCCCTATC	AGATAATGAC
351	CGGGCTGTTG	CTGCTGTATA	TGCTGGCGAT	GCGGTTTGTG	TTGCAAGAAG

401 CGCGCGCCAA AACCGACTTC CGGCACATTG CCGTCTCGCG CGCGGTTGTC
 451 GCGGAGCCG GCTATTTTCG CGGCCATTG AGTTACTACG ACCGGGTGCG
 501 GATGGCAAT ATCTTCTGCGC CAACAACATT CTACTACGCC AAGAATCAGG
 551 CGATGCTCTA CACCGTCAGC CAGAATGCG AGTTTATTAC CGCGCGCTTG
 601 GTGATCCCG TCTTCTCTCC CTTGGGCAAT CAACAGGCTG CGCGCCAGCA
 651 TCTGAACGAG CGGAATCTC AAAAAATCCT CTTTATCGTC GCCGAATCTT
 701 GGGGGCTGCC GGCCAAATCC GAATCTCAAA ACGCCACTTT TGCCAAACTG
 751 TCTGGGCAAA AAGACCGTTT TTCCGTTTGG GAAAGCGGCA GTTTTCCCTT
 801 CATCGGCGCG ACGGTCGAG CGGAAATGCG CGAATCTGTG GCCTACGGCG
 851 GTTTCGCGCG GTTCGCACTG CGCGCGCGCG CGACAGAAAA ATTTGCCCGC
 901 TGCTCTCCCA ACCGTTTGAA ACAAGAAGGT TAGCCCACTT TTGCGATGCA
 951 CGCGCGCGCG AGTTCTGCTT ACACACGCTT CAGCTGGTAT CGAGGCGCGG
 1001 GCTTTCAAGA AATCAAAACC GCGCAAAACC TGATCGGTAA AAAAACCTCG
 1051 GCAATTTTCG CGCGCGGTGT CGACAGCGAG CTGTTTCGGC AAGTGTCCG
 1101 ATTTTTCAAA AAACACGACA AGGGACTGTT TTAAGTGGAT ACGTGAACCA
 1151 GCGACGCGCA CTATCCGAAA TCGGACATT TCAACACAG GCTCAATATG
 1201 ACGGAATATG GCCTGCCGCG CGAAACCGAC CTCGCGCGCA ATTTCAAGCT
 1251 GCACACCCAA TTCTCTGACG AACTGCGGGA TTGATCTCAA CGCCCGGAAA
 1301 TGAAGGCAC GGAAGTCACT ATCGTCGGCG ACCATCGCGC GCGCTCGCG
 1351 AACTCTAATG AAACCTTCCG CTACTCTAAA CAGGSGCAGC TCGCTGGCT
 1401 GAATCTCAA ATCAATAA

This corresponds to the amino acid sequence <SEQ ID 470; ORF48-1>:

1 MNHTLLSKQ WLPFFLEPR LLSLLILLA PNAVFWLAL LTATARPIVN
 51 LDYLPALLI ALPWRFKIA GVLAFWLAVL FDGLMMVIO FPFMDLIGAI
 101 NLVFFILTAP APYQINTGLL LLYLAMPEV LQKAAKTD F RHIAVCAAVV
 151 AAAGYFTGHL SYYDRGRMAN IFGANNFYIA KQAMLYTVS QNADFTAGL
 201 VDPVFLPLGN QORAAATHLE FKSQKILFV AESWGLPANP ELQNAFTAKL
 251 LAQKDRFSVW ESGSPFFIGA TVEGEMRELK AYGLRLGFAL RRAPEKDFAR
 301 CLPNRLKQEG YATFAMHAGS SLYDRFSWY PRAGEQIKT AENLIGKTK
 351 AIFGGVCDSE LFGEVSAPFK KHDKGLFYWM TLTSHADYPE SDIFNRLKLC
 401 TEYGLPAETD LCRNFSLHTQ FFDQADLIQ RPEMKGTVEI IVGDHPFVPG
 451 NLNETRYLK QGVHAWLNFK IK*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

35 ORF48 shows 94.1% identity over a 119aa overlap with an ORF (ORF48a) from strain A of *N. meningitidis*:

	10	20	30	40	50	60
orf48 pep	MNIHTLLSKQWTLPPFLPKRLLLSLLILLAPNAVFWLALLTATARPIVNLDPYLAALLI					
orf48a	MNIHTLLSKQWTLPPFLPKRLLLSLLILLXPNVFWLALLTATARPIVNLXYLAALLI					
	10	20	30	40	50	60
	70	80	90	100	110	119
orf48 pep	ALPWRFKIAGVLAFWLAVLFDGLMMVIOFPPFMDLIGAINLVFFILTAPAPYQINTGL					
orf48a	ALPWRFKIXGVLAFWLAVLFDGLMMVIOFPPFMDLIGAINLVFFIXTAPALYQINTGLL					
	70	80	90	100	110	120
orf48a	LLYLAMPEVFLQKAAKTDERHIAACAAVVVAAGYFTGHLXSYDRGRNANIFGANNFYIA					
	130	140	150	160	170	180

The complete length ORF48a nucleotide sequence <SEQ ID 471> is:

1 ATGAATATTC ACACCTGCTC CTCCAAACAA TGGACGCTG CGCATTCCTC
 51 GCGGAACCGG CTGCTGCTGT CCCTGCTGAT ACTGCTNNCC CCGAATGCGG
 101 TGTTTTGGGT TTTGGCAGTG CTGACGCGCA CGCGCCGCCG GATTGTCAAT
 151 TTGGANTACC TTCCGCGCGC GCTGCTGATC GCCCTGCCTT GCGGTNTCGT
 201 CAAAATGNC GCGCTATTGG CGTNTTGCTT GCGCGTTTGT TTTGACGGCG
 251 TGATGATGGT GATCCAACTC TTCCCTTTTA TGGATCTCAT CGCGCCATC
 301 AACCTCGTCC CTTTTCATCT GACCGCCCCC GCGCTTTATC AGATAATGAC
 351 CGGCTGTGTA CTGCTGTATA TGCTGGCGAT GCGCTTTGTG TTGCAGAAAG
 401 CGCGCCGCAA AACCGACTTC CGACACATTG CGCGCTGTGC CGCGGTTGTG
 451 GTGGCAGCGG GCTATTTTAC CGGCCATTG AGTTANTACG ACCGGGGGGG

501 GATGGCCART ATCTTCGGCG CAACACACTT CTATTACGCC AAAAGTCAGG
 551 CGATGCTCTA CACCGTCAGC CAGAATGCCG ACCTTATTAC CGCGCGCGTG
 601 GTCGATCCCG TCTTCCTCCC CTTGGGCAAT CAACAGCGTG CGCGCCACGA
 651 TCTGAACGAG CGGAATCTC AAAAAATCCT CTTTATCGTC GCCGAATCTT
 701 GGGGGCTGCC GGCCAATCCC GAACTTCAAA ACGCCACTTT TGCCAAACTG
 751 CTGGCGCAAA AAGANCGTTT TCGGTTTGG GAAAGCGGCA GTTTTCCTTT
 801 CATCGGCGCG ACGATCGAAG GCGAATGCG CGAAGCTGTG GCCTACGGCG
 851 GTTTGCGCGG GTTCGCACTG CGCGCGCGCG CGGACGAAA ATTTCGCCGC
 901 TGCTCCCCCA ACCGTTTGAA ACAAGAAGGT TACGCCACTT TTGCGATGCA
 951 CGGCGCGCGG AGTTGCGTTT ACGACCGTT CAGCTGGTAT CCGAGGGCGG
 1001 GCTTTCAAGA AATCAAAACC GCGCAAAAC TGATCGGTAA AAAAACCCTG
 1051 GCAATTTTCG CGCGCGTGTC CGACAGCGAG CTGTTCCGGC AAGTGTCCGC
 1101 ANTTTTCAAA AAACACGACA AGGGACTGTT TTAAGTGATG ACGCTGACCA
 1151 GGCACGCGCA CTATCCGGA TONGACATT TCAACACAG GCTCAAAATG
 1201 ACCGAATATG GCTTCGCCG CGAAACCGAC NTCTCGCGCA ATTTACAGCT
 1251 GCACACCCAA TCTTCGACC AACTGGCGGA TTGATCCAA CGCCCGAAA
 1301 TGAAAGGCAC GGAAGTCATC ATCTGCGCGC AACCATCGCG GCCGCTCGG
 1351 AACCTCAATG AAACCTCCG CTACCTCAA CAGGGGCGAG TCGNTGGCT
 1401 GAACCTCAA ATCAATAA

20 This encodes a protein having amino acid sequence <SEQ ID 472>:

1 MNHITLLSKQ WLPPEFLPKR LLSLLILLX PNAVFWLAL LTATARPVFN
 51 LXPVPAALLI ALPWRXVKIX GVLAAXLAVL FDGLMMVQL PFMDLIGAI
 101 NLVPFXTAP ALYQIMTGLL LLYMLAMPFV LQKAAKTDF RHIAACAAV
 151 VAAGYFTGHL SKYDRGRMAN IFGANNFYA KSQAMLYTS QNADFITAGL
 201 VDPVFLPLGN QRAATHLNE PKSQKILFV AESWGLPANP ELQNATFAKL
 251 LAQKXRFVSU ESGSFFFIGA TIEGEMREL CAYGGLRGFAL RRAPDEKFA
 301 CLPNRLKQEG YATFAMHAGG SLYDRFSWY PRAGFQEIKT AENLIGKKT
 351 AIFGGVCDSE LFGEVSAXFK KHKDGLFWM TLTSHADYPE SDIFNHRLEK
 401 TEYGLPAETD XCRNFSLHTQ FFDQLADLIQ RPENKGTVEI IVGDHPFVPV
 451 NLNETRYLKL QGHVXWLNFK IK*

ORF48a and ORF48-1 show 96.8% identity in 472 aa overlap:

		10	20	30	40	50	60
35	orf48a.pep	MNHITLLSKQWLPPEFLPKRLLSLLILLX	PNAVFWLALLTATARPVFN	LNKLYLPAALLI			
	orf48-1	MNHITLLSKQWLPPEFLPKRLLSLLILLX	PNAVFWLALLTATARPVFN	LNKLYLPAALLI			
		10	20	30	40	50	60
40	orf48a.pep	ALPWRXVKIXGVLAAXLAVLFDGLMMVQL	PFMDLIGAINLVPFXTAP	ALYQIMTGLL			
	orf48-1	ALPWRXVKIXGVLAAXLAVLFDGLMMVQL	PFMDLIGAINLVPFXTAP	ALYQIMTGLL			
		70	80	90	100	110	120
45	orf48a.pep	LLYMLAMPFVLQKAAKTDFRHIAACAAVV	VAAGYFTGHL	SKYDRGRMAN	IFGANNFYA		
	orf48-1	LLYMLAMPFVLQKAAKTDFRHIAACAAVV	VAAGYFTGHL	SKYDRGRMAN	IFGANNFYA		
		130	140	150	160	170	180
50	orf48a.pep	KSQAMLYTSQNADFITAGLVDPVFLPLGN	QRAATHLNEPKSQKILFV	AEWSGLPANP	ELQNATFAKL		
	orf48-1	KSQAMLYTSQNADFITAGLVDPVFLPLGN	QRAATHLNEPKSQKILFV	AEWSGLPANP	ELQNATFAKL		
		190	200	210	220	230	240
55	orf48a.pep	ELQNATFAKL	LAQKXRFVSU	ESGSFFFIGA	TIEGEMREL	CAYGGLRGFAL	RRAPDEKFA
	orf48-1	ELQNATFAKL	LAQKXRFVSU	ESGSFFFIGA	TIEGEMREL	CAYGGLRGFAL	RRAPDEKFA
		250	260	270	280	290	300
60	orf48a.pep	CLPNRLKQEGYATFAMHAGG	SSLYDRFSWY	PRAGFQEIKT	AENLIGKKT	CAIFGGVCDSE	
	orf48-1	CLPNRLKQEGYATFAMHAGG	SSLYDRFSWY	PRAGFQEIKT	AENLIGKKT	CAIFGGVCDSE	
		310	320	330	340	350	360
65	orf48a.pep	CLPNRLKQEGYATFAMHAGG	SSLYDRFSWY	PRAGFQEIKT	AENLIGKKT	CAIFGGVCDSE	
	orf48-1	CLPNRLKQEGYATFAMHAGG	SSLYDRFSWY	PRAGFQEIKT	AENLIGKKT	CAIFGGVCDSE	

-280-

		370	380	390	400	410	420
orf48a.pep		LFGEVSAFFKKHDKGLFYWMTLTSHADYPESDIFNHRILKCTEYGLPAETDXCRNFSLHTC					
5 orf48-1		LFGEVSAFFKKHDKGLFYWMTLTSHADYPESDIFNHRILKCTEYGLPAETDLNCRNFSLHTC					
		370	380	390	400	410	420
orf48a.pep		FFDCLADLILQRPEMKGTEVIIVGDHPVPVGNLNETFRYLKQGHVXALNFKIKX					
10 orf48-1		FFDCLADLILQRPEMKGTEVIIVGDHPVPVGNLNETFRYLKQGHVXALNFKIKX					
		430	440	450	460	470	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF48 shows 97.5% identity over a 119aa overlap with a predicted ORF (ORF48ng) from *N.*

gonorrhoeae:

orf48.pep	MNIHTLLSKQWTLPPFLPKRLLSLLILLAPNAVFWLALLTATARPVNLVDYLPALLI	60
orf48ng	MNIHALLSEQWTLPPFLPKRLLSLLILLAPNAVFWLALLTATARPVNLVDYLPALLI	60
20 orf48.pep	ALPWRFKVIAGVLAFLAVLFDGLMMVIQLFPFMDLIGAINLVFFILTAPAPYQIMTGL	119
orf48ng	ALPWRFKVIAGVLAFLAVLFDGLMMVIQLFPFMDLIGAINLVFFILTAPAPYQIMTGL	120

The ORF48ng nucleotide sequence <SEQ ID 473> was predicted to encode a protein having amino acid sequence <SEQ ID 474>:

25	1	MNIHALLSEQ WTLPPFLPKR LLSLLILLAP NAVFWLAL LTATARPVNL
	51	LDYLPALLI ALPWRFKVIA GVLAFLAVL FDGLMMVIOL FPFMDLIGAI
	101	NLVFFILTAP APYQIMTGLL LLYLAMFV LQAAVKTDF RHIAVCAAVV
	151	AAARYFTGPF ELLRTGGRWQ YVQHRLLLS GSRASFRRRQ KADVLRRLGN
	201	PYASMGNGG..

Further work identified the complete gonococcal DNA sequence <SEQ ID 475>:

	1	ATGAATATTC	ACGCCCTGCT	CTCCGAACAA	TGGACGCTGC	CGCCATTCCT
	51	CGCGAAACGG	CTGCTCGTGT	CCCTGCTGAT	ACTGCTGGCC	CCCAATCGGG
	101	TGTTTTGGGT	TTTGGCACTC	CTGACGCCA	CGCGCGCGC	GATTTGCAAT
	151	TTGGACTACC	TTCCGCGCGC	GCTGCTGATC	GCGCTGCGCT	GCGGTTTGCT
35	201	CAAAATTGCC	GCGCATTTGG	CGTTTGGCC	GCGGCTTTTG	TTTGACGGGC
	251	TGATGATGGT	GATCCAATC	TTCCCTTTTA	TGGACCTCAT	CGCGGCCATC
	301	AACTCTGCTC	CTTCTCATCT	GACCGGCCCC	GCGCCTTATC	AGATAATGAC
	351	CGGGCTGTTG	CTGCTGTATA	TGCTGGCGAT	GCGGTTTGTT	TTGCAAAAAG
	401	CGCGCGTCAA	AACCGACTTC	GCACACATTG	CGGCTGTGCG	CGCGTTTGTT
40	451	GCGGCGACCG	GCTATTTCAC	CGGCAATTG	AGTTACTACG	ACCGGGGGCG
	501	GATGGCCAAAT	ATCTTCGGCG	CAACAACACT	CTATTACGCC	AAAGTCAGG
	551	CGATGCTCTA	CACCGTCAGC	CAGAATCCG	ACTTTATTAC	CGCCGgcttc
	601	GTGACACCCG	TCTTCTCCCT	CTTGGCAAT	CAGCAGCGTG	CGCGACCGGG
45	651	GCTGAGTGAG	CGAAATCTC	AAAAAATCCT	CTTTATCGTC	GCGCAATCTT
	701	GGGGGCTGCC	GGGCAATCCC	GAGCTTCAAA	ACGCCCACTT	TGCCAACTG
	751	CTGGGCGCAA	AAGACGGTTT	TTGCGTTTGG	GAAAGGGGCA	GTTTTCCTTT
	801	CATCGGCGCG	ACGCTGGAAG	CGGAAATGCG	GGAATTGTGC	GCTACGGGG
	851	GTTTGGCGCG	GTTGCGACTG	CGCGCGCGCG	CGGACGAAAA	ATTTCGCGCG
	901	TGCTTCCCA	ACGCTTTGAA	ACGAGAAGGT	TACGCTGCTT	TTGCGATGCA
50	951	CGCGCGGGT	ACTTTCGCTT	ACGACGCGTT	CAGCTGCTAT	CGGAGCGCG
	1001	GCTTTCAAAA	AATCAAAACC	CGCGAAAACC	TGATCGGTAA	AAAAACCTGC
	1051	GCCATTTTCG	GCGGCGTGTG	CGACAGCGAG	CTGTTGGGCG	AAGTCTCGCG
	1101	ATTTTTCAAA	AAACACGACA	AGGGACGTGT	TTAATGGATG	ACGCTGACCA
	1151	GCCACGCGGA	CTATCCCGAA	TCCGACATT	TCAACACACG	GCTCAATATG
55	1201	ACCGAATACG	GCGTCCCGCG	CGAAACCGAC	CTCTCGCGCA	ATTTCAAGCT
	1251	GCAACCCCAA	TtcttcgACC	AACTGGCGGA	TTTGATCCGA	CGCCCGCGAA
	1301	TGAAAGGCAC	GGAAGTCATC	ATCGTCCGCG	ACCATCCGCG	GCCGCTCGCG
	1351	AACTCTCAAT	AAACCTTCGG	CTACTCAAAA	CAGGACACAG	TGCGCTCGCT
	1401	GCACTTCAAA	ATCAATAATA			

This encodes a protein having amino acid sequence <SEQ ID 476; ORF48ng-1>:

```

1  MNHALLSEQ WTLPPFPLPKR LLLSLILLIA ENAVFVVLAL LTATARPIVN
51 LDYLEAALLI ALPWRFKIA GVLAWFVVL FDGLMVFQI FFMDLIGAI
101 NLVPEITAP ARYQIMTGLL LLYMLAMPFV LQKAAVKTDF RHIAVCAAVL
151 AAGYFTGHL SYDRGRMAN IFGANNFYIA KSQAMLYTVS QNADFTTAGL
201 VDPVFLPLGN QORAATRLSE PKSOKILFTV AESWGILGNP ELQNATFAKL
251 LAQKDRFSVW ESGSFPTGA TVEGEMREL CAYGGLRGLF RRAPEKFAF
301 CLPNRLKQEG YATFAMHGAG SSSLYDRFSWY PRAGFOEIKT AENLIGKKTCAI
351 AIFGGVCDSE LFGEVSAFFK KHDGKLFYWM TLTSHADYPE SDIFNHLKTC
401 TEYGLPAETD LCRNFSLHTQ FFDQLADLIR RPEMKGTEVI IVGDHPPFVG
451 NLNETFRYLK QGHVAVLHFK IK*

```

ORG48ng-1 and ORF48-1 show 97.9% identity in 472 aa overlap:

```

10 20 30 40 50 60
orf48-1.pep MNHITLLSKOWTLPPFPLPKRLLLSLLILLAPNAVFWVLALLTATARPIVNDYLPAAALI
15 orf48ng-1 MNHITLLSKOWTLPPFPLPKRLLLSLLILLAPNAVFWVLALLTATARPIVNDYLPAAALI
10 20 30 40 50 60
70 80 90 100 110 120
20 orf48-1.pep ALPWRFKIAGVLAFWLAVLFDGLMVFQIQLFPFMDLIGAINLVPFILITAPAPYQIMTGLL
orf48ng-1 ALPWRFKIAGVLAFWLAVLFDGLMVFQIQLFPFMDLIGAINLVPFILITAPAPYQIMTGLL
70 80 90 100 110 120
25 orf48-1.pep LLYMLAMPFVLQKAAKTDFRHIAVCAAVVAAAGYFTGHLSSYDRGRMANIFGANNFYIA
orf48ng-1 LLYMLAMPFVLQKAAKTDFRHIAVCAAVVAAAGYFTGHLSSYDRGRMANIFGANNFYIA
130 140 150 160 170 180
30 orf48-1.pep KSOAMLYTVSONADFITAGLVDPVFLPLGNQORAATHNEPKSOKILFTIVAESWGLPANP
orf48ng-1 KSOAMLYTVSONADFITAGLVDPVFLPLGNQORAATHNEPKSOKILFTIVAESWGLPGNP
190 200 210 220 230 240
35 orf48-1.pep ELQNATFAKLLAOKDRFSVWESGSPFFIGATVEGEMRELCAYGGLRGLFALRRAPDEKFAF
orf48ng-1 ELQNATFAKLLAOKDRFSVWESGSPFFIGATVEGEMRELCAYGGLRGLFALRRAPDEKFAF
250 260 270 280 290 300
40 orf48-1.pep CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFOEIKTAENLIGKKTCAIFGGVCDSE
orf48ng-1 CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFOEIKTAENLIGKKTCAIFGGVCDSE
310 320 330 340 350 360
50 orf48-1.pep LFGEVSAFFKKHDKGLFYWMTLTSHADYPESDIFNHLKCTEYGLPAETDLCRNFSLHTQ
orf48ng-1 LFGEVSAFFKKHDKGLFYWMTLTSHADYPESDIFNHLKCTEYGLPAETDLCRNFSLHTQ
370 380 390 400 410 420
55 orf48-1.pep FFDQLADLIRPEMKGTEVIIVGDHPPVGVNLTFRYLKQGHVAVLHFKIKX
orf48ng-1 FFDQLADLIRPEMKGTEVIIVGDHPPVGVNLTFRYLKQGHVAVLHFKIKX
430 440 450 460 470

```

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and two putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, can be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 57

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 477>:

```

1  ..GTGAGCGGAC GTTACCGGCG TTTGGATCGC GTTTCCAAAA TCATCATCGT
51  TACTTTTGAST ATCGCCACGC TTGCGCGCGC CGGCAATCGT ATGTCCGCGG
101  GTATCGAGAT GCAGTCCGAT TTTATCGAGC CGACACCGTG GACGCTTGGC
151  GCTTTGGGCT TCGTATCGGC GCTGATGGCG TGGATCGCGC CGCGCATTTA
201  AATTTCGCGC ATCAATTCTT TGTGGTAAC CGAAAACAA CGCATCAATC
251  CTTCGGAATA CCGCGACGGG ATTTTGAAT TCAACGTCGG TTATATCGCC
301  AGTGCGGTTT TGGCTTTGGT TTTCTTGCA CTGGCGCG.G TAGCGCGAA
10  351  CGGCAACGGC GA.ACAGTGC AGATGGCGGG CGGCAATAT AACGGGCAAT
401  TGATCAATAT GTACGCG..

```

This corresponds to the amino acid sequence <SEQ ID 478; ORF53>:

```

1  ..VSGRYRALDR VSKIIIVTLS IATLAAAGIA MSRGMQMSD FIEPTPTWLA
51  GLGLIALMNG WMPAFIEISA INSLWVTEKQ RINPSEYRDG IFEENVGYIA
15  101  SAVLALVFLA LGXVAPNNG XTVQAGGKY NQOLINMYA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 479>:

```

1  ATGTCCGAAC AACATATTTT GACTTGGAAA AGTAAAATCA ACGCATTGGG
51  TCGCGGGATC ATGATGGGCT CGCGCGGGGT CGCGGGTTCG CACCTGATTG
101  CCTCGACGCA GGGCGGGCGG CTTTACGGCT GGCAGATCGC GCTCATCATC
20  151  ATCCTGACCA ACCTCTTCAA ATACCGGCTT TTCCGCTTCA GCGCGATTAA
201  CACGCTGGAC ACGGCAAGA GCGTGAATGA AGSTATTGCC GAGAAAAGCC
251  GCGTTTATT GTGGGATATC CTGATTTTGT GCATCTCTTC CGCCAGCATG
301  AACCGGGGCG CGGTGCGCAT TGTAAACGCG GCATCTCTCA AATATGCGAT
351  TCCCTCGGCT ATCTTTGATG CCGCGCGGTT TCGCGCTTGT ATTATGGCAT
25  401  CCGCGCTGAT TATTTGGTGT AGCGGACGTT ACCCGGCTTT GGATCGCGTT
451  TCCAAAATCA TCATCGTTAC TTTGAGTATC GCCACGCTTG CGCGCGCGG
501  CATCGCTATG TCGCGGGGTA TGCAGATGCA GTCCGATTTT ATCGAGCCGA
551  CACCGCTGGAC GCTTGGCGGT TTGGGCTTCC TGATCGCGCT GATGGGCTGG
601  ATGCCCGGCG CGATTTGAAAT TTCCGCCATC AATTCTTTGT GGGTAACCGA
30  651  AAAACAACGC ATCAATCCTT CGGAATACCG CGACGGGATT TTTGATTTCA
701  ACGTCGGTTA TATCGCCAGT GGGGTTTGG CTTTGGTTT CTCTGCACTG
751  GCGCGGTTTG TGCAATACGG CAACGGCGAA CGAGTCGAGA TGGCGGGCGG
801  CAAATATATC GGGCAATTGA TCAATATGTA CGCGTTTACC ATCGCGCGCT
851  GGTGCGGGCC GCTGGTGGCG TTTATCGGCT TTGCTGTAT GTACGCGACG
35  901  ACGATTACCG TCGTGGACGG CTATGCCGCT GCATATGCGG AACCGGTGGG
951  CTTGCTGGCG GGAAGAAGCA AAACGGGCAA CGCGAATTC TTTGCTGCG
1001  ATATTGGGTT GCGGGGCGAG GGTTTGGCGG TGATTTTCTG GTTTGACGGC
1051  GTAATGGCGA ATCTGGCTCA ATTCGCGATG ATTGCGGCTT TGTGTCGCGC
1101  CCTGCTGTTT CTTGCTGTA ATTCGCTT GTTAAAGGT GATGAAGAC
40  1151  ACAACCTCAC ATCAGGTATG ATGCGCTT GATTGCGAGG CTGATTATT
1201  CTGACCGGTT TTACCGTTTT GTTCTTATTG AATTTGGCGG GAATGTTCAA
1251  ATGA

```

This corresponds to the amino acid sequence <SEQ ID 480; ORF53-1>:

```

1  MSEQHSITWK SKINALGPI MMASAAVGGG HLIASTQAGA LYGWQIALII
45  51  ILNLFKYPF FRFSAHYTLT TKGSLIEGYA EKSRYVLWVF LILCILSATI
101  NAGAVAVTVA AIVKMAIPLS MFDAGTVAAL IMASCLILLV SGRYRALDRV
151  SKIIIVTLSI ATLAAAGIAM SRGMQMSDF IEPTPTWLAG LGFLIALMNG
201  MPAPFIESAI NSLWVTEKOR INPSEYRDGI FDNVGYIAS AVLALVFLAL
251  GAFVQYNGGE AVQMAGGKYI GQLINMYAVT IGGWSRPLVA FIAFACMYGT
50  301  TITVVDGYAR AIAEPVRLLR GRDKTGNAEF FAWHINWAGS GLAVIEWFDG
351  VMANLLKFM IAIFYSAFVF ANLNYRLVKG DEKHLTSGM NALALAGLIY
401  LTGFTVLFLI NLAGMFK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

55 ORF53 shows 93.5% identity over a 139aa overlap with an ORF (ORF53a) from strain A of *N. meningitidis*:

```

      10          20          30
orf53.pep    VSGRYRALDRVSKIIIVTFLSIATIAAGIA
              |||||
5 orf53a      AAIVKMAIPSLMFDAGTVAAALIMASCLIIIVSGRYRALDRVSKIIIVTFLSIATIAAGIA
             110       120       130       140       150       160
              |||||
orf53.pep    MSRGMQMOSDFIEPTFWTLAAGLFIALMGWMPAPIEISAINSLWTEKQRINPSEYRDG
              |||||
10 orf53a     MSRGMQMOSDFIEPTFWTLAAGLFIALMGWMPAPIEISAINSLWTEKQRINPSEYRDG
             170       180       190       200       210       220
              |||||
orf53.pep    TFEINVGYIASAVIALVFALGXVPANGXTQVMAGGKYNGQLINMYA
              |||||
15 orf53a     IFDENVGYIASAVIALVFALGAIFYOYNGEAVQMAGGKYTGOLINMYAVTTIGGSRLPV_
             230       240       250       260       270       280
              |||||
20 orf53a     AFIAFACMYGGTITVVVDGARAIAPVLLRGKDKTNAEFFAWNIIWAGSGLAVIFWD
             290       300       310       320       330       340

```

	1	ATGTCGCAAC	AACATATCT	GACTTGGAA	AGTAAAGTA	ACGCATTGGG
	51	ACCGGGGAT	ATGATGGCTT	GCGGGGGGAT	CGGGGGTTGC	CACTGATTGT
25	101	CTCTGCAGCA	GCGGGGGGCG	CTTTACGGCT	GCGAGATCGC	GCTCATCATC
	151	ATCTCGACCA	ACCTCTTACA	CTATCCGTTA	TTCCGGCTTA	GCGGCGATTA
	201	CACGCTGCAC	ACGGGACAGA	GCGTTAGTGA	AGGTTATGCC	GAGAAAAGAT
	251	CGGTTTATTT	GTGGGGTTCT	CTGATTTTGA	GACCTCTCTC	CGCACGATT
	301	AACCGGGGCG	CGGTGCGCAT	TGTAAACGCC	GCCATCTGCA	AAATGGCGAT
	351	TCCTCTGTG	ATGTTTGTAT	CGCGGACGAT	TGCGGCGCTT	ATTATGGCTT
30	401	CTCGCTGAT	TATTTTGTGT	AGCGGACATC	ACCGCGCTTT	GGATCGGCTT
	451	TCGAAATAT	TACGCTGAT	TTTGGATGTT	GCGAGCGTCT	CGGCGCGGGT
	501	CACTCGGGCT	TCGAGCGGTA	TCGAGCGGTA	TCGAGCGGTA	TCGAGCGGTA
	551	CACCGTGCAC	GCTGCGCGGT	TTGGCGTTCT	TGATCGCGCT	GATGGGCGCT
	601	ATGCGCGGCG	CGATTGAAAT	TCCGCGCATC	AATCTTTGTT	GGGTAAACGA
35	651	AAACACACGC	ATCAATCTTT	CCGAAATCGC	GCACGGGATT	TTTGATTTCA
	701	ACGTGGGTTA	TATCGCCGAT	GCGGTTTTCG	TTTGCTTTTT	CTTTCGACTG
	751	CGAGCGTTTG	TGCAATACGG	CACGCGGACG	GCAGTGCAGA	TGGCGGGGCT
	801	CAATATATCT	GCGCAATATG	TCAATATGTA	CGCGCTTACC	ATCGCGGGCT
40	851	GGTGCGGCCG	GCTGGTGGCG	TTTATCGGCT	TGCGCTGTAT	GTAACGGGCT
	901	AGGATACCG	TTGTGGACGG	CTATGCGGT	GCCATTCGCG	ATACCGGTGC
	951	CTCGCTGCGC	GGAAGAAAGC	GAACCGGACG	CGCGCGAATC	TTTTCGTTGA
	1001	ATATTTTGGT	GCGGGCGAGC	GGTTTGGCGG	TGATTTTCTG	TTTTCGAGCG
	1051	GTAATGGGCA	ATCTGCTGAT	ATTTCGCGAT	ATGCGCGGCT	TGTTGCTGGT
45	1101	CGCTGCTGTA	ATGAGGCGTA	ATGAGGCGTA	ATGAGGCGTA	ATGAGGCGTA
	1151	ACAACATCAC	ATCACTGTAT	AATGCGCTCT	CTTCTGACG	CTTGATTTAT
	1201	CTGACCGGTT	TTACCGTTTT	GTTCTTATTT	AATTTGCGCG	GAAATGTCAA
	1251		ATGA			

50	1	MSEOHISITWK	SKINALGPI	MMASAVOGGS	HLIATSTAGS	LYWQWIALII
	51	ITLNLFLKVF	FRFSAHYTL	TKSLGIVGA	KSRVYUWLF	LILCSILATI
	101	NGAGVAIVTA	AIVKMAISL	MEDAGTVAFL	IMASCLIIIL	SGRYALDRV
	151	SKI IIVTSL	ATLAAAGASL	SRFGMOSDF	IEPTPTWLAG	LGLFIALIMG
	201	MAPFELISAI	NSLWVTKR	INPEYRDTI	FDNNGVIA	AVIALFVLAL
	251	GFAPYOGS	AVOMAGQRI	GOLINNYAT	IGWGRSLVA	FIAFACMYG
55	301	TTIVVDYAT	IASPRLVGL	KROKTNAGS	FAMNWTGSG	GLAVIFWPG
	351	WVANLISL	IAFVSAPAF	AMNLTALVK	DEKHLITAG	NALALGJLI
	401	LTGATGTL	NIAGMEK*			

60 crf53a.pep MSEQHISTWKSINALGPGINMASAAVGGSHLIASTQAGALYGWQTALIIILTNLFKYFF
|||
crf53-1 MSEQHISTWKSINALGPGINMASAAVGGSHLIASTQAGALYGWQTALIIILTNLFKYFF
|||
10 20 30 40 50 60
65 70 80 90 100 110 120

	orf53a.pep	FRFSAHYTLDTGKSLIEGYAEKSRVYLWVFLILCILSATINAGAVAIVTAAIVKMAIPSL	
	orf53-1	FRFSAHYTLDTGKSLIEGYAEKSRVYLWVFLILCILSATINAGAVAIVTAAIVKMAIPSL	
5		70 80 90 100 110 120	
	orf53a.pep	130 140 150 160 170 180	
10	orf53-1	130 140 150 160 170 180	
	orf53a.pep	190 200 210 220 230 240	
15	orf53-1	190 200 210 220 230 240	
	orf53a.pep	250 260 270 280 290 300	
20	orf53-1	250 260 270 280 290 300	
	orf53a.pep	310 320 330 340 350 360	
25	orf53-1	310 320 330 340 350 360	
	orf53a.pep	370 380 390 400 410	
30	orf53-1	370 380 390 400 410	
35		370 380 390 400 410	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF53 shows 92.1% identity over a 139aa overlap with a predicted ORF (ORF53ng) from *N.*

gonorrhoeae:

40	orf53.pep	VSGRYRALDRVSKIIIVTLSIATLAAAGIA	30
	orf53ng	AAIVKMAIPSLMFDAGTVAALINASCLIIIVSGRYRALDRVSKIIIVTLSIATLAAAGIA	91
45	orf53.pep	NSRGMQMOSDFIEPTFWTLAGLGLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDG	90
	orf53ng	NSRGMQMOSDFIEPTFWTLAGLGLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDG	151
	orf53.pep	IEFENVGYIASAVLALVFLALGXVAPNGXGTVMAGGKYNGLINNYA	139
50	orf53ng	IFDENVGYIASAVLALVFLALGXVQYNGEAVQMGGGKYGQILINNYAVTIGGSRPLV	211

An ORF53ng nucleotide sequence <SEQ ID 483> was predicted to encode a protein having amino acid sequence <SEQ ID 484>:

51	1	MPKSCVYLW VFLILCIASA TINAGAVAIV TAAIVKMAIP SLMFDAGTVA
51	51	ALIMASCLII IVSGRYRALD RVSKIIIVTL SIATLAAAGI AMSRGMQNP
101	101	DFIEPTFWTL AGLGLIALM GWMPAPIEIS AINSWVTEK QRINPSEYRD
151	151	GIFDENVGYI ASAVLALVEL ALGAFVQYNG GEAVQMGGGK YIGQLINNYA
201	201	VTIGGSRPLV VAFIAFACMY GAASTVVDGY ARAIAEPVRL LRGDKTARF
251	251	IVLLEKIGGR HRFGRDFLV*

Further analysis revealed further partial DNA gonococcal sequence <SEQ ID 485>:

60	1	..aagaAAGACT GCGTTTATTT GTGGGTTTTT TTGATTITGT GTATCGCCTC
51	51	CGCCACGATT AACGGGGGG CGGTGCGCAT TGTAAACGCC GCCATCGCTA
101	101	AAATGGCGAT TCCTCGCTG ATGTTTGATG CCGCACGGT TGCGGCTTG


```

5      251  ATTTATGGCAT  CCTGCCTGAT  TATTTTGGTG  AGCGGAAGTT  ACCGCGCTTT
      261  GGATCGGTGT  TCCAAATCA  TCATTGTTAC  TTGAGCATCG  GCCAGCTGTG
      271  CCGCGCGCGG  CATCGCTATG  TCGCGCGGTA  TCGAGATGCA  GCCCGATTTT
      301  ATCGAGCGGA  CACCGTGGAC  GCTTGCCTGT  TTGGGCTTCC  TSGATCGCGT
      351  GATGGGCTGG  ATGCCGCGCG  CGATCGAAAT  TTCCGCGATC  AATTCTTTGT
      401  GGGTAACCGA  AAAACAACGC  ATCAATCCTT  CTGAATACCG  CGACGGGATT
      451  TTGATTATCA  ACGTCCGGTA  TATCGCagT  GCGGTTTGGT  CTTTGGTTTT
      501  CCTTGCACGT  GGGCGGTTTG  TGCATAACGG  CAACGGCGAA  GCAGTGAGA
      551  TGGCGGGGCG  CAAATATATC  GGGCAATTGA  TTAATATGTA  TGCCGTAAAC
10     601  ATCGGCGGCT  GGTCTGCTCC  GCTGGTGGCG  TTATCGCGGT  TTGCGCTGTAT
      651  GTACGCGCAG  ACGATTACCG  TTGTGGACGG  TTATGCGCGT  GCCATTGCGC
      701  AAACCGTGGC  CCTGCTGGCG  GGCAGGGATA  AAACCGGCAA  CGCOGAGTTG
      751  TttgccTGGa  ATATTGGGT  GGGGGGCGAG  GGTTTGGCGG  TGATTTTCTG
      801  GTTTGACGgc  gcaaTGGCg  AACTgcTCAA  ATTTGOGATG  AttgcccgcCT
15     851  TTGTGTCGCG  CCTGTGTTTC  GCTTGGCTCA  ACTACCGCCT  CGTCAAAGGG
      901  GACRAACGCC  ACAGGCTTAC  GCGCGGTATG  AACGCCCTTG  CCATTGTGCG
      951  CTTGCTCTAC  CTGGCGGGT  TTGCGTGTTC  GTTCCTGTG  AACCTTACCG
1001   GACTTTTGGC  ATAG

```

This corresponds to the amino acid sequence <SEQ ID 486; ORF53ng->:

```

20      1  ..KSCVYLWVF  LILCIASATI  NAGAVAIVTA  AIVKMAIPSL  MFDAGTVAAL
      51  IMASCLIIIV  SGRYRALDRV  SKIIIVTISI  ATLAAGIAM  SRGMQMOPDF
      101  IEPTFWTLAG  LGFLIALMGW  MPAPETISAI  NSLWVTEKQR  INPSEYRDGI
      151  FDFNVGYIAS  AVLALVFLAL  GAFVQYNGNE  AVCMAGGKYI  GQLINMYAVT
25     201  IGGWSRPLVA  FIAFACMYGT  TITVVDGYAR  AIAEPVRLLR  GRKGTGNAEL
      251  FAWNINWAGS  GLAVIFWFDG  AMAELLKFAM  IAAFVSAPVF  AWLNRYRLVK
      301  DKRHRLTAGM  NALAIVGLLY  LAGFAVLFL  NLTGLLA*

```

ORF53ng-1 and ORF53-1 show 94.0% identity in 336 aa overlap:

```

30      60      70      80      90      100      110
orf53-1.pep  ILTNLFKYPPFRFSAHYTLDTGKSLIEGYAEKSRVYLWVFLILCILSATINAGAVAIVTA
orf53ng-1    :|| ||||| ||||| ||||| ||||| |||||
              KKSCVYLWVFLILCIASATINAGAVAIVTA
              10      20      30

35      120      130      140      150      160      170
orf53-1.pep  AIVKMAIPSLMFDAGTVAALIMASCLIIIVSGRYRALDRVSKIIIVTSLIATLAAGIAM
orf53ng-1    AIVKMAIPSLMFDAGTVAALIMASCLIIIVSGRYRALDRVSKIIIVTSLIATLAAGIAM
              40      50      60      70      80      90

40      180      190      200      210      220      230
orf53-1.pep  SRGMQMOSDFIEPTFWTLAGLGFLIALMGWMPAPETISAINSLWVTEKQRINPSEYRDGI
orf53ng-1    SRGMQMOPDFIEPTFWTLAGLGFLIALMGWMPAPETISAINSLWVTEKQRINPSEYRDGI
              100      110      120      130      140      150

45      240      250      260      270      280      290
orf53-1.pep  FDFNVGYIASAVLALVFLALGAFVQYNGNEAVCMAGGKYIGQLINMYAVTIGGWSRPLVA
orf53ng-1    FDFNVGYIASAVLALVFLALGAFVQYNGNEAVCMAGGKYIGQLINMYAVTIGGWSRPLVA
              160      170      180      190      200      210

50      300      310      320      330      340      350
orf53-1.pep  FIAFACMYGTTITVVDGYARAIAEPVRLLRGRKGTGNAEFFAWNINWAGSGLAVIFWFDG
orf53ng-1    FIAFACMYGTTITVVDGYARAIAEPVRLLRGRKGTGNAELFAWNINWAGSGLAVIFWFDG
              220      230      240      250      260      270

60      360      370      380      390      400      410
orf53-1.pep  VMANLLKFAMIAAFVSAPVFAWLNRYRLVKGDKEHKLTSGMNALALAGLIYLTGFTVLFL
orf53ng-1    AMAELLKFAMIAAFVSAPVFAWLNRYRLVKGDKRHRLTAGMNALAIVGLLYLAGFAVLFL
              280      290      300      310      320      330

65      orf53-1.pep  NLAGMFKX
              ||:|:|
      orf53ng-1    NLTGLLAX

```

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 58

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 487>:

```

1    ..TTGCGGGAAA CGGCATATGT TTTGGATAGT TTTGATCGTT ATTTTGTGTG
51   TCGCGTTGCC GGCTTGT TTT TGTCCGCGC ACAATCCGAA CGCAGATGGA
101  TCGCGCAGGT TTTGCGGTGG CAGGAAAGA AAGGGGAAAA ACAGCGGAG
151  CTGCCTGAAA TCAAGAAGCG TATGCCGAT TTTCCCGAAC TTGCCCTGAT
201  GCTTTTCCAC GCCGTCAAAA CGGCAGTGTA TTGCTGT TTT GTCGGTGTGC
251  TCCGTTTCTG CCGAACTAT TTTGCCGACG AATCGAACC GGACAGGCC
301  GTTCCGCCT..

```

15 This corresponds to the amino acid sequence <SEQ ID 488; ORF58>:

```

1    ..LRETAYVLDS FDYFVVALA GLFFVRAQSE REWMREVS AW QEKKGKQAE
51   LPEIKDGMFD FPEALMLFH AVKTAVYVLF VGVVRFRCRNY LAHESEPD RP
101  VFP..

```

Further work revealed the complete nucleotide sequence <SEQ ID 489>:

```

20  1    ATGTTTGTGA TAGTTTTGAT CGTTATTTTG TTGCTTGCGC TTGCGCGGTT
51   GTTTTTTGTG CGCGCACAAAT CGGAAACGCGA GTGGATGCGC GAGGTTTCTG
101  CGTGCGCAGGA AAGAAGAAGGG GAAAAACAGG CGGAGCTGCC TGAATACAAA
151  GACGCGATATG CGGATTTTCC CGAATCTGCC CTGATGCTTT TCCATGCCGT
201  CAAAACGCGCA GTGTATTGCG TGTATTGTCG TGTCGCTCGT TCTCGCGAA
251  ACTATCTGCG GCACGATCC GAAACGCGAC GCGCGCTTC GCTGCTCTCT
301  GCAAACCGTG CGGATGTTCG GACCGCATCC GACGATTTT CAGAAGTGG
351  AAACGGGACG GAAGAAGCGG AAACGGAAGA AGCAGAGCTT GCGGAGGAG
401  AGGCTGCGGA TACGGAAGAC ATTGCAACTG CGTAAATCGA CAACGCCCGC
451  ATCCCATTCG ACCGGAATAT TGCTGAAGGG TTGATGCCGT CTGAAAGCGA
501  AATTTCGCCG GTCCGTCGCG TTTTAAAGA AATCACTTTG GAAGAAGCAA
551  CGCGTGCTTT AAACAGCGCG GCTTTAAGGG AAACGAAAAA ACGCTATATC
601  GATGCAATTTG AGAAAAACGA AACAGCGGTC CCCAAGTCC GCGTGCCGA
651  TACCCGATG GAAGGGCTGC AGATTATCGG TTTGACGAC CCGTGCTTC
701  AACGCAAGTA TTCCCATATG TTGATGCGG ACAAGAAGCG GTTTCCGAG
751  TCTGCGGATT ACGGATTGGA GCGTATTTT GAGAAGACG ATCCGCTCTG
801  CTTTCTTGCA GTCAAGACCG AATATGCAAG GAATGCGCG TTCCACCGTG
851  ATGCAAGGCA GGGGAAAGGG CAGGCGGAGG CAAAATCCCC GGATGTTTCC
901  CAGGGGCACT CGGTTTCAAG CCGCACGCGC GTCCGCGATG CCGCGCGCGC
951  CGTTTCCGCT AATTGAAAG AACCGAAACAA GGCAACGGTT TCTGCGGAG
1001 CGGAAATTTT TCGCTCGATT CCGGAAAGTC AGACGGTTGT GCGGAAACGG
1051 GATGCGGAAA TCGCTGTGTA AACCGAAATG GTTTTCAGCG AAACCGTTTC
1101 GTCTGTGGGA TACGCGGCTC CGGTTGCGCG TGAACTGCGC GATATCATA
1151 TTGAAGAACC TGCGCGGCCG GATGCTTGGG TGTCGAAC ACCCGAAGTG
1201 CGGAAGTTC CCATGACGCG AATCGATATT CAGCGCGCGC CTCCGCTATC
1251 GGAATCTAC AACCGTACCT ATGAAACCGC GTCAAGTATC GAGCAGGTGC
1301 AACGCGCGCG CATTTGCCAG ACCGACCATC TTGCGGATGA TGTTTTGAAT
1351 GGAAGTTGGC AGGAGGAAC CGCGCTATT CGGATGACG CGAGTGAAGG
1401 TCGCGCAGAG CGGTCAAGCG GGCAATATCT GTGCGAAACC GAAGCGTTTC
1451 GGCATGACAG TCAGGCGGTT TGTCGCTTTG AAATGTGTC GTCTGAACCG
1501 CCGTCTGCGC GGGTATCGGA TACGGAAGCG GATGAAGGGG CGTTCCCATC
1551 TGAAGAAACC GGTGCGGTAT CGGAACACCT CGCGACAACC GACCTGCTTC
1601 TGCCCTCGCT GTTCAATCCC GAGGCGACGC AAACGGAAGA AGAAGTGTG
1651 GAAACAGCA TCACCATCGA AGAAAAATTG CGGAGTTTCA AAGTCAAGGT
1701 CAAGGTTGTC GATTCTTATT CCGGCCCCGT AATTACGGGT TATGAATTCG
1751 AACCGGATGT CGGCGTGCGC GGCAATTTCG TCTGAATCT GGAAAGATG
1801 TTGGCGCGGT CGCTCGGCGT GGCTTCCATC CGGCTGTGTC AAACCATCCC
1851 CGGCAAAACC TGCATGGGTT TGGAACTTTC

```

1901	TACGCCTGAG	CGAAATCTTC	AAITTCGCCG	AGTTTGGCGA	ATCCAAATCC
1951	AAGCTGACCC	TCCGCTCGG	TCCAGACATC	ACCGCAGC	CCGTGGTAC
2001	CGACTTGGGA	AAAGCACCGC	ATTTTGTTGGT	TGCCCGCAGC	ACCGGCTTCG
2051	GCAATCGGT	GGGTGTCAAC	CGGATGATTC	TGCTATGCT	TTTCAAGACC
2101	GGCCGGAAG	ACGTGGGTAT	GATTATGATC	GATCCGAAAA	TGCTGGAATT
2151	GAGCATTAC	GAAGGCATCC	CGCACTGCT	CGCCCTCTGC	GTTACCGATA
2201	TGAAGCTGGC	GGCAACCGCG	CTGAACCTGGT	GTGTTAACGA	AATGGAAAAA
2251	CGCTACCGCC	TGATGAGCTT	TATGGGCGTG	CGTAATCTTG	CGGCTTCAA
2301	TCAAAAATC	GCCGAGCGCG	CAGCAAGGGG	AGAAAAATC	GGCAATCCGT
2351	TCAGCCTCAC	GCCCGACGAT	CCCGAACCTT	TGGAANAATC	GCGCTTTATC
2401	GTGGTGTGG	TGATGAGT	TGCCGACCTG	ATGATGACGG	CAGGCAGAAA
2451	AATCGAAGAA	CTGATTGCOC	CGCTCGCCCA	AAAAGCCCGC	CGGCGAGGCA
2501	TCCATTTGAT	TCTTGCCACA	CAACGCCCCA	CGCTCGATGT	CATCACGGGT
2551	CTGATTAAAG	GCAACATCC	GACGCGTATC	GCGTCCCAAG	TGTCACGAAA
2601	AATCGACAGC	CGACGATTC	TGCACCAAT	GGGCGCGGAA	AACTGCTCG
2651	GTACGGCGGA	TATGCTGTTT	CTGCTGCCGG	GTACTGCCCTA	TCCGACGCGC
2701	GTTCCAGCGC	CGTTGCGCTC	GGATGAAGAG	GTCGACCGCG	TGCTCGAATA
2751	TTTGAAACAG	TTTGGCGAAC	CGGACTATGT	TGACGATATT	TGAGCGCGCG
2801	CGGCGAGCGA	AGAGCTGCC	GGCATCGCG	CGAGCGGCGA	CGACGAATCC
2851	GATCCGATGT	ACGACGAGGC	CGTATCGGT	GTCCTGAAAA	CGCGCAAAAC
2901	CAGCATTTCC	GGCGTACAGC	GGCGCTGCG	TATCGGCTAC	AACCGCGCGC
2951	CGCGTCTGAT	TGACCAGATG	GAGCGGAGAG	GCATTGTGCT	CGCACCGGAA
3001	CACAAACGCA	ACCGTACGAT	TCTCGTCCOC	TTGACAAATG	CTTGA

This corresponds to the amino acid sequence <SEQ ID 490; ORF58-1>:

25	1	MFIVLIVIL	LLALAGLFFV	RAQSEREMNR	EVSAAQEKKG	EKQALPEIK
	51	DGMPDFPELA	LMFLHAVKTA	VYVLFVGVVR	FCRNYLAHES	EPDRPVFPAS
	101	ANRADVPPTAS	DGYSDSNGMT	EEAEETEEAEA	AEAEAAADTED	IATAVIDNR
	151	IPFDRSIABG	LMPSESEISP	VRPVKEITL	EEATRALNSA	ALRETKKRYI
	201	DAFEKNETAV	PKVRVSDTPM	EGLQIIGLDD	PVLQRTYSHM	FDADKEAFSE
30	251	SADYGFPEPYF	EKQHPSAFSA	VKAENARNAP	PHRHAGQGCG	QAEAKSPDVS
	301	QQGVSDGTA	VRDARRRVSV	NLKEPNKATV	SAEARI SRLI	PESQTVVGR
	351	DVEMPESETEN	VETETVSVSG	YGGPVYDETA	DIHIEEPAAP	DAWVVEPPEV
	401	PKVPMTAIDI	QPPPVSEIY	NRTEPPSPG	EQVRSRIAE	TDHLADDVLN
35	451	GGWQETAAI	ADDSGEGAAE	RSSGQLSET	EAFGHSQAV	CFPENVPSE
	501	PSCRVSTETA	DGAFSESET	GAVSEHLPTT	DLLLPLFNF	ENPTQEEELL
	551	EMSTIEELK	AEFFVKVKIV	DSVSGPVITR	VELEPDPVVR	GNVNLNED
	601	LARSLGVASI	RVVETTPGKT	CMGLEPLNPK	RQMIRLSEI	NSPEFAESKS
	651	KITLALGQDI	TGQPVVTDLG	KAPHLVAGT	TSGSKSVGVN	AMILSMLFKA
40	701	APEDVRMIMI	DFKMLESIIY	EGIPHLAPV	VTDMKLAANA	LNWCNMEMEK
	751	RYRLNSFMGV	RNLGAFNQKI	AEAAARGEKI	GNPFSLTPDD	PEPLEKLPI
	801	VVVVDEFADL	MMTAGKIEE	LIRALQAQAR	AAGIHILAT	QRPVSDVITG
	851	LIKANIPTRI	AFQVSSKIDS	RTILDQMGAE	NLLGGQDMFL	LLPGTAYPQR
	901	VHGAFASDEE	VHRVVEYLKQ	FGEPTYVDI	LSGGGSEELP	GIGRSGDDET
	951	DPMYDEAVSV	VLKTRKASIS	GVQRALRIGY	NRAARLIDQM	EAEQVSAFE
45	1001	HGNRTILVP	LDNA*			

Computer analysis of this amino acid sequence predicts the indicated transmembrane region, and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF58 shows 96.6% identity over a 89aa overlap with an ORF (ORF58a) from strain A of *N.*

meningitidis:

		10	20	30	40	50	60
orf58.pep	LRETAYVLDSD	FRYFVVALAGLFFV	RAQSEREMNR	EVSAAQEKKG	EKQALPEIK	DKGMPD	
orf58a		MFIVLIVILL	LLALAGLFFV	RAQSEREMNR	EVSAAQEKKG	EKQALPEIK	DKGMPD
			10	20	30	40	50
orf58.pep	FP	ELALMLFHAVKTA	VYVLFVGVVR	FCRNYLAHES	EPDRPVFP		
orf58a	FP	ELALMLFHAVKTA	VYVLFVGVVR	FCRNYLAHES	EPDRPVFP	PASANRADVP	TASDGYSD
		60	70	80	90	100	110

The complete length ORF58a nucleotide sequence <SEQ ID 491> is:

```

1  ATGTTTGGGA TAGTTTGGAT CGTTATTTTG TTGCTTGCGC TTGCCGGCTT
51  GTTTTGTGTC CGCGCCACAAT CGGAACGCGA GTGGATGCGC GAGGTTTCTG
101  CGTGGCAGGA AAGAGAAAGG GAAAAACAG CGGAGCTGCG TGAAATCAAA
151  GACGGTATGC CGGATTTTCC CGAACTTGCC CTGATGCTGTT TCCATGCCGT
201  CAAAAACGGA GTGTATTGGC TGTTTGTCGG TGTCGTCGCT TTCTGCGGAA
251  ACTATCTGGC GCACGAATCC GAACGCGACA GGCCCGTTCC GCCTGCTCTC
301  GCAAAATCGTG CGGATGTTCC GACCCGATCC GACGGATATT CAGACAGATGG
351  AAGCGGACGC GAGAGACCGG AAGACGACGT ACGGAGAGAG CGGAGAGAG
401  AGGCTSCCGA TAGCGAAGAC ATTGCAACTG CGTAATCGGA CAAACGCCGC
451  ATCCCATTCG ACCGGAGTAT TGTCTGAAGG TTGATGCGCT CTGAAGCGCA
501  AATTTCGCCCC GTCCGTCGGG TTTTAAAGGA AATCACTTTG GAAGAAGCAA
551  CGCGTGCCTT AACACGCGCG GCTTTAAGGG AAGCGAAAAA ACGCTATATC
601  GATGCATTATG AGRAAAACGA AACAGCGGTC CCCAAGTCC CGGTGTCGGA
651  TACCCCGGATG AAGGGGCTCG AGATTATCGG TTGAGCAGAC CCGTGTGCTTC
701  AACGCACGATA TTCCCGTATG TTGATGCGGG ACAGAAGAGC GTTTTCCGAG
751  TCTGCGGATT ACGGATTGGA GCCGTATTTT GAGAAGCAGC ATCCGCTGTC
801  CTTTCTGCGA GTCAGAGCGC AATATGCACG GAATGCGCGC TTCCGCGCTG
851  ATGCAGGGCA GGGNAAAGGG CAGGCGGAGG CNAATCCCCC GGATGTTTCC
901  CAAGGGCAGT CCGTTTCACA CGGCACAGCC GTCCGCGGAT CCNCGCGCGC
951  CGTTTCCGCT AATTGAAAG AACCGAACA GGCACAGGTT TCTGCGGAGG
1001  CGCGGATTTC GCGCCTGATT CGGGAAGTCC GGACGGTTGT CGGGAAACGG
1051  GATGTCGAAA TGCCGCTCTGA AACCGAAAAA GTTTTTCAGG AAANTGTTTC
1101  GTCCTGTTGGA TACCGCGCTTC CGGTTTATGA TGAACACTGCC GATATCCATA
1151  TTGAAGAAAC TCCGCCGCTG GATGCTGGG TTGCTGAAC CCCTGAAATG
1201  CGGAAGATTC CCAATGCCCGC AATGATATT CCGCCGCGC CTCCGATATC
1251  GGAATCTTAC AACCTTACTC ATGAACCGCC GGCAGGATTC GAGCAGGTG
1301  AACGCAAGCG CATTGCCGAA ACCGATCATC TTGCCGATGA TGTTTGAAT
1351  GGAGGTTGGC AGGAGGAACG CGCGCTATT GCGAATGACG CAGGTGAGGG
1401  TGTGGCAGAG CGGTCAAGCG GGCATATATT GTCGGAACCC GAAGCGTTCG
1451  GGCATGACAG TCAGGCGGTT TGTCCGTTTG AAAATGTGCC GTCTGAACGC
1501  CCGTCCCGCG GGGCATNGGA TACGGAAGCG GATGAAGGGG GGTCCAATC
1551  TGAAGAAACC GGTGCGGTAT CCGAACACCT CGCGCAACCC GACCTGCTTC
1601  TGCCGCGCGT GTTCAATCC GGGGCGACGC AAGCGGAAGA AGANCTGTTG
1651  GANAACAGCA TCACCATCGA AGAAAAATNG GCGGAGTTCA AAGTCAAGGT
1701  CAAGGTGTTC GATTCTTATT CGCGCCCGGT GATTACGCGT TATGAATGCT
1751  AACCCGATGT CGCGCTGCGC GGCATTCGCG TTCTAAATCT GGAAGAAAGN
1801  TTGGCGCGTT CGCTCGCGGT GGCCTCCATC CGCGTGTGCG AAACACTCTC
1851  CGGCAAAACC TGATATGGTT TGAACACTCC GAACCCGAAA CGCCAAATGA
1901  TAGCCCTGAG GAAATCTTC AATTGCGCGG AGTTTGCGGA ATCCAAATCC
1951  AAGCTGACGC TCCGCTCGCG TCAGGACATC ACCCGACAGC CCCTGTAAC
2001  CGCTCTGGCG AAGCAGCCGC ATTGTTGGT TCCGCGCAGC ACCGTTGCGG
2051  GCAAAATCGT GGGTGTCAAC GCGATGATC TGCTATGCT TTTCAAGGCC
2101  CGCGCGGAAG ACCTGCGTAT GATTATGATC GATCCGAAAA TGCTGGAATT
2151  GAGCATTTAC GAAGCATCC CGCACCTGCT CGCCCTGTCT GTTACCGATA
2201  TGAAGCTGGC GGCACACGCG CTGAACCTGT GTGTTAAAGA AATGGAAJAA
2251  CGCTACCGCG TGATGAGCTT TATGGCGGTG CGCAATCTTG CGGCTNTCAA
2301  TCAAAARATC GCGCAAGCGC CAGCAAGGGG GGAGAAATCT GGCAACCCGT
2351  TCAGCTCTAC GCGCGACAAT CCGCAACCTT TGGANAATTT CGCGTTTATC
2401  GTGGTCTGTT TTGATGAGTT TGCGCACTGT ATGATGACGG CAGGCAAGAA
2451  AATCGAAGAA CTGATGTCOC GCTCGCCCA AAAAGCCGCG CGCGCAGGCA
2501  TCCATCTTAT CTTTGCCACA CAACGCCCCA GTGTCGATGT CATCACGGGT
2551  CTGATTAAAG CGAATCATCC GACGCGTATC CGGTTCCAGG TGTCAGGACA
2601  AATCGACAGC CGCAGGATTC TTGAACAAAT GGGTGGCGGA AACCTGCTCG
2651  GGCAGGGCGA TATGCTGTTC TCGCGCGCGG GTACGCGCTA TCGCGAGCGC
2701  GTTCAGCGCG GTTTGTGCTC GGATGAAGAG GTGACGCGCG TGGTGGAATA
2751  TCTGAACACG TTTGGCCGAC CGGACTATGT TGCGCATATN TTGAGCGCGG
2801  GTATGTCGCA GATTTGCTG GAAATCAGCG GAGCGCGGCA CGCGAARACC
2851  GATCCGATGT ACGAGCGAGC CGGTGCTGCT GTTTTGAAGA CCGCGAARAC
2901  CAGCATTTCT GCGCTGCGAC GCGCATTTGG TATGCGCTAT AATCGCGCGC
2951  CGCGCTCTGT TGACGAGATG GAGGCGGAAG GCATTGTGTC CGCACCGGAA
3001  CACAAACGCA ACCGTACGAT TCTGCTCCCC TTNGACAATG CTTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 492>:

```

1  MFWIVLIVIL LLALAGLFFV RAQSEREWMR EVSAWQEKKG EKQALPEIK
51  DGMFPDFPELA LMLFHAVKTA VYWLFGVVFR FCNRYLAHES EPDRPVPPAS
101  ANRADVPVTA DGYSDSGNGT EEAETEAEAE AEEEAEDTDE IATAVIDNRR
151  IPRFDSIAEG IMPSESEISF VRVFKETL EEAATRALNS ALRETKKRYI
201  DAFEKNETAV PKVRVSDTDM EGLQIIGLDD PVLQRTYSRM FDAKKEARSE

```

251	SADYGFEPYF	EKQHPSAFSA	VKAENARNAP	FRRHAGQGGK	QAEAKSPDVS
301	QGQSVSDGTA	VRDAXRRVSV	NLKEPNKATV	SAEARISRLI	PESRTVVGKR
351	DVEMPSETEN	VFTEKXVSSVG	YGPVYDETA	DIHIEEPAAP	WDAMVVEPPEV
401	PKVMPFAXDI	PPPPVPVSEIY	NRTYEPPAGF	EQVQSRSAIE	TDHLADDVLN
451	GGWQEETAAI	ANDGSEGVAE	RSSGQVLSSET	EAFGHDSQAV	CFPENVPVSR
501	PSRRAXDTEA	DEGAFQSEET	GAVSEHLPTT	DLPLPFLNP	GATQTEEXLL
551	XNSITIEEKK	AEFKVKVQV	DSYSGPVITR	YEIEPDVGV	GNSVLNLEKK
601	LARSLGVASI	RVVETILGKT	CMGLELPNPK	ROMIRLSEIF	NSPEFAESKS
651	KLTALGQDI	TGQPVVTDLG	KAPHLLVAGT	TGSGSKSVGVN	AMILSMLEKA
701	APEDVRMIMI	DPKMLELSIY	EGIPHLLAPV	VTDMKLAANA	LNNCWVMEKE
751	RYRLMSFMGV	RNLAGXNKIK	AEAAARGEKI	GNPFSLTDPN	PEPLKCLFTI
801	VVVVDEFAFL	MMTAGKKIEE	LIARLAQKAR	AAGIHLILAT	QRPSVDVITG
851	LKANIPTRI	AFQVSSKIDS	RTILDQMGAE	NLLGQGDMLF	LPPGTAYPQR
901	VHGAFASDEE	VHRVVEYLKQ	FGEFDYVDDK	LSGGMSSDLL	GISRSQDGET
951	DEMYDEAVSV	VLKTRKASIS	GVQKRLRIGY	NRAARLIDQM	EAEQIVSAPE
1001	HNGNRTILVP	KDNA*			

ORF58a and ORF58-1 show 96.6% identity in 1014 aa overlap:

		10	20	30	40	50	60
20	orf58a.pep	MFWIVLIVILLALAGLFFVRAQSEREMMREVSANQEKGEKQAELEIKDKGMPDFPELA					
	orf58-1	MFWIVLIVILLALAGLFFVRAQSEREMMREVSANQEKGEKQAELEIKDKGMPDFPELA					
		10	20	30	40	50	60
25	orf58a.pep	LMLFHAVKTAVYWLFGVVRFCRNYLAHESEFDRFPVPPASANRADVPTASDGYSDSGNGT					
	orf58-1	LMLFHAVKTAVYWLFGVVRFCRNYLAHESEFDRFPVPPASANRADVPTASDGYSDSGNGT					
		70	80	90	100	110	120
30	orf58a.pep	EEAEETEAEAEAEAAEDTEDIATAVIDNRRIPFDRSIAEGLMPSESEISFVRPVKEITL					
	orf58-1	EEAEETEAEAEAEAAEDTEDIATAVIDNRRIPFDRSIAEGLMPSESEISFVRPVKEITL					
		130	140	150	160	170	180
35	orf58a.pep	EEATRALNSAALRETKKRYIDAFENKETAVPKVRVSDTPEGLQITGLDDPVLQRTYSRM					
	orf58-1	EEATRALNSAALRETKKRYIDAFENKETAVPKVRVSDTPEGLQITGLDDPVLQRTYSRM					
		190	200	210	220	230	240
40	orf58a.pep	EEATRALNSAALRETKKRYIDAFENKETAVPKVRVSDTPEGLQITGLDDPVLQRTYSRM					
	orf58-1	EEATRALNSAALRETKKRYIDAFENKETAVPKVRVSDTPEGLQITGLDDPVLQRTYSRM					
		190	200	210	220	230	240
		250	260	270	280	290	300
45	orf58a.pep	FDADKEAFSESADYGFEPYFEKQHPSAFSAVKAENARNAPFRRHAGQGGKQAEAKSPDVS					
	orf58-1	FDADKEAFSESADYGFEPYFEKQHPSAFSAVKAENARNAPFRRHAGQGGKQAEAKSPDVS					
		250	260	270	280	290	300
50	orf58a.pep	QGQSVSDGTAVRDAXRRVSVNLEPNKATVSAEARISRLIPESRTVVGKRDVEMPSETEN					
	orf58-1	QGQSVSDGTAVRDAXRRVSVNLEPNKATVSAEARISRLIPESRTVVGKRDVEMPSETEN					
		310	320	330	340	350	360
55	orf58a.pep	VFTEKXVSSVGYPVYDETAADIHIEEPAAPDAWVVEPPEVKVMPFAXDI PPPVPVSEIY					
	orf58-1	VFTETVSSVGYPVYDETAADIHIEEPAAPDAWVVEPPEVKVMPFAXDI PPPVPVSEIY					
		370	380	390	400	410	420
60	orf58a.pep	NRTYEPPAGFEQVQSRSAIETDHLADDVLNGGWQEETAIAINDGSEGAERSSGGYLSET					
	orf58-1	NRTYEPPAGFEQVQSRSAIETDHLADDVLNGGWQEETAIAINDGSEGAERSSGGYLSET					
		430	440	450	460	470	480
65	orf58a.pep	EAFGHDSQAVCFPENVPSESRPSRRAXUTEADEGAFQSEETGAVSEHLPTD LLLPPLFNP					
	orf58-1	EAFGHDSQAVCFPENVPSESRPSRRAXUTEADEGAFQSEETGAVSEHLPTD LLLPPLFNP					
		490	500	510	520	530	540
70	orf58a.pep	EAFGHDSQAVCFPENVPSESRPSRRAXUTEADEGAFQSEETGAVSEHLPTD LLLPPLFNP					
	orf58-1	EAFGHDSQAVCFPENVPSESRPSRRAXUTEADEGAFQSEETGAVSEHLPTD LLLPPLFNP					
		490	500	510	520	530	540

		550	560	570	580	590	600
	orf58a.pep	GATQTEEXLLXNSITIEEKXAEFKVKVVDSSYSGPVITRYEIEPVDVGVRGNSVLNLEKX					
5	orf58-1	EATQTEELLNSITIEEKLAEFKVKVVDSSYSGPVITRYEIEPVDVGVRGNSVLNLEKX					
		550	560	570	580	590	600
	orf58a.pep	LARSLGVASIRVVETILGKTCMGLPNPKRQMRILSEIFNSPEFAESKSKLTALGQDI					
10	orf58-1	LARSLGVASIRVVETIPGKTCMGLPNPKRQMRILSEIFNSPEFAESKSKLTALGQDI					
		610	620	630	640	650	660
	orf58a.pep	TGQPVVTDLGKAPHLLVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIY					
15	orf58-1	TGQPVVTDLGKAPHLLVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIY					
		670	680	690	700	710	720
	orf58a.pep	EGIPHLLAPVVTDMKLAANALNWCNEMEKRYRLMSFMGVNRLAGXNKIAEAAARGEKI					
20	orf58-1	EGIPHLLAPVVTDMKLAANALNWCNEMEKRYRLMSFMGVNRLAGXNKIAEAAARGEKI					
		730	740	750	760	770	780
	orf58a.pep	GNPFSLTNDPNEPLXKLPIVVVDFADLMNTAGKIEELIARLAQKARAAGIHLILAT					
25	orf58-1	GNPFSLTDDPEPLEKLPIVVVDFADLMNTAGKIEELIARLAQKARAAGIHLILAT					
		790	800	810	820	830	840
	orf58a.pep	QRPSVDVITGLIKANIPTRIAFOVSSKIDSRITLDQMAENLLQGQDMLFLPGTAYPQR					
30	orf58-1	QRPSVDVITGLIKANIPTRIAFOVSSKIDSRITLDQMAENLLQGQDMLFLPGTAYPQR					
		850	860	870	880	890	900
	orf58a.pep	VHGAFASDEEVHRVVEYLKQFGPEPDYVDDXLSGGMSDDLGLISRSQDGETDPMYDEAVSV					
35	orf58-1	VHGAFASDEEVHRVVEYLKQFGPEPDYVDDXLSGGMSDDLGLISRSQDGETDPMYDEAVSV					
		910	920	930	940	950	960
	orf58a.pep	VHKTRKASISGVORALRIGYNRAARLIDQMEAEIGVSAPEHNGNRTILVFXDNAX					
40	orf58-1	VHKTRKASISGVORALRIGYNRAARLIDQMEAEIGVSAPEHNGNRTILVFXDNAX					
		970	980	990	1000	1010	
	orf58a.pep	ALMLFHAVKTAIVYWLFGVVRFCRNYLAHESEPDVPVPP					
45	orf58-1	SEPDRVPVPASANRADVPTASDGYSDSGNG					

50 Homology with a predicted ORF from *N.gonorrhoeae*

ORF58 shows complete identity over a 9aa overlap with a predicted ORF (ORF58ng) from *N. gonorrhoeae*:

	orf58.pep	ALMLFHAVKTAIVYWLFGVVRFCRNYLAHESEPDVPVPP	103
55	orf58ng	SEPDRVPVPASANRADVPTASDGYSDSGNG	30

The ORF58ng nucleotide sequence <SEQ ID 493> is predicted to encode a protein having partial amino acid sequence <SEQ ID 494>:

	1	..SEPDRVPVPPA	SANRADVPTA	SDGYSDSGNG	TEEAETEAAE	AAEEEAADTE
60	51	DIATAVIDNR	RIPFDRSIAE	GLMQSESKTS	PVRPVKEIT	LEEARLASS
	101	AALRTEKKRY	IDAFERKNGTA	VPKVRVSDTP	MEGLQIGLD	DPVLQRTYSR
	151	MFDADEKAFS	ESADYGFEYP	FEKQHPSAFS	AVKAENARNA	PFRRHAGQEK
	201	GQAEAKSPDV	SQQQSVDGDT	AVRDARRRVS	VNLKEPNKAT	VSAEARISRL
	251	IFESRTVVGK	RDVEMFSET	NVFTETVSSV	GYGGPVYDEA	ADIHIEEPPA
	301	PDAAWVEPPE	VPEVAPEID	ILPPPVSEI	YNRTYEPFAG	FEQAQRSRIA

351	ETDHLAADVL NGGWQEETAA IADGSEGA ERSSGGYLSE TEAFGHDSQA
401	VCFFEDVPSE RPSCRVSDTE ADEGAFQSEE TGAVSEHLPT TDLLEPPLFN
451	REAVTTEEL LENSITIEEK LAEFKVRNV VDSISGPVIT RVEIEDVGV
501	RGNVTLNEK DLARSLGVAS IRVETIPGK TCMGLELNP KRQMLRLSEI
551	FNSPFEAEK SKITLALGQD ITGQPVVTDL GKAPHLLVAG TTGSGKSVGV
601	NAMILSMLEK AAPEDVRMIM IDPKMLELSI YEGITHLLAP VVTDMKLAAN
651	ALNWCVNEME KRYRLMSFGV VRNLAGFNOK IAEAAARGEK IGNPFLTPD
701	DPEPLEKLFP IVVVDEFAD LMMTAGKKE ELIARLAQKA RAAGIHLILA
751	TQRPSPDVIT GLIKANIPTR IAFQVSSKID SRTILQDMA ENLLGGQMDL
801	FLPPTGATPQ RVHGAFASDE EVHRVVEYLK QFGEPDYVD ILSGGSEEL
851	PGIRSGDGE TDPYDEAVS VVLKTRKASI SGVQRLAIRG YNRAARLIDQ
901	NEAEGIVSAP EHNGNRTILV PLDNA*

This partial gonococcal sequence contains a predicted transmembrane region and a predicted ATP/GTP-binding site motif A (P-loop; double underlined). Furthermore, it has a domain homologous to the FTSK cell division protein of *E. coli*. Alignment of ORF58ng and FtsK (accession number p46889) show a 65 % amino acid identity in 459 overlap:

ORF58ng:	467	IEEKLAEFKVKVKVSDVSYSGPVITRYEIEPDVGRGNSVLNLEKDLARSLGVASIRVVET	526
FtsK:	868	+E +LA+P++K VV+ GPVITR+E+ GV+ + NL +DLARSL ++RVVE	
ORF58ng:	527	IPGKTCMGLELNPVKRQMLRLSEIFNSPEFAESKSKITLALGQDITGQPVVTDLGKAPHL	586
FtsK:	928	IPGK +GLELPN KRQ + L E+ ++ +F ++ S LT+ LG+DI G+PVP DL K PHL	
ORF58ng:	587	LVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIYEGITHLLAPVVTDMK	646
FtsK:	988	LVAGTTGSGKSVGVNAMILSML+KA PEDVR IMIDPKMLELS+YEGI HLL VVTDMK	
ORF58ng:	647	LAANALNWCVNEMEKRYRLMSFGVVRNLAGFNOKIAEAAARGEKIGNPESLTPDDPEF--	704
FtsK:	1048	DAANALWCVNEME+RY+LMS +GVNRNLAG+M+KLAE I +P: D +	
ORF58ng:	705	--LEKLPFIVVVDEFADLMMTAGKKEELIARLAQKARAAGIHLILATORPSVDVITGL	762
FtsK:	1108	L+K P+IVV+VDEFADLMMT GKK+ELIARLAQKARAAGIHL+LATORPSVDVITGL	
ORF58ng:	763	IKANIPTRIAFOVSSKIDSRITLDQMAENLLGGQMDLFPFGTAYPQRVHGAFASDEEV	822
FtsK:	1168	IKANIPTRIAF VSSKIDSRITLDQ GAE+LLG GDL+ P + P RVHGA F D+EV	
ORF58ng:	823	HRVVEYLKQFGEPDYVDILSGGSEELPGIRSGDGETDPYDEAVS VVLKTRKASISG	882
FtsK:	1228	HVV+ K G P YVD I S SE G G G E DP++D+AV V + RKASISG	
ORF58ng:	883	VQRLRIGYNRAARLIDQMEAEIGVSAPEHNGNRTILV 921	
FtsK:	1287	VOR RIGYNRAAR+I+QMEAR+GIVS HNGNR +L P	
		1287 VQRFRIGYNRAARIIQMEAQGIVSEQHNGNREVLAP 1325	

Further work on ORF58ng revealed the complete gonococcal DNA sequence to be <SEQ ID 495>:

1	ATGCTTTGGGA TAGCTTTTGAT CGTATTgtg TTGCTTGCCT TGCOCGGCCT
50	51 GTTCTTTTGTC CGCGACAAT CGGAACGCGA GTGATGCGC GAGGTTCTCTG
101	CGTGCGACGA AAGAAGAGGG GAAAAACAGG CGGAGCTGCC TGAATACAA
151	GACGGTATGC CCGATTFTTCC CGAGTTTTC CTGATGCTTT TCGATGCGGT
201	CAAAACGCGA GTGATTATGGC TGTTTGTGCG TGTCGTCGCT TCTGCGCGAA
251	ACTATCTGCG GCACGAATCC GAACCGGACA GGCCCGTTC CCGCTGCTTCT
301	GCAAAACGCGT CGGATGTTC GACCGCATCC GAOCGTATT CAGACAGTGG
351	AACCGGGAGC GAAGAAGCGG AACCGGAAGC AGCAGAAGCT GCGGAGGAAG
401	AGGCTGCGCA TACGGAAGAC ATTGCAACTG CGTAATCGA CAACCGCGCG
451	ATCCcatTCG ACCGAGTAT TGCTGAAGGG TTGATGCAGT CTGAAGAGCA
501	AACCTGCGCC GTCCGTCGG TTTTAAAGGA AATCACTTTG GAAGAAGCAA
551	CGCGTGCTTT AAGCAGCGCG GCTTTAAGGG AAGCGAAAA ACGTATATC
601	GATGCATTTC AGAAAAACGG AACAGCCGTC CCCAAAGTAC CGGTGTCGGA
651	TACCCCGATG GAAGGCGTCG ACATTATCGG TTGGACGAC CCGTGTCTTC
701	ACCGCAGTA TCCCGGTATG TTGATGCGGG ACAAGAGAG GTTTCGCGG
751	TCGCGGATT ACAGATTGGA GCCGTATTTT GAGAGCAGC ATCCGCTGTCG

801	CTTTCTGCA	GTCAAAGCG	AAAATGCACG	GAATGCGCG	TTCCGCGTC
851	ATGCAGGCA	GSAGAAAGGG	CAGCGCGAGC	CAAAATCCCG	GGATGTTTTC
901	CAGGCACT	CGTTTCGCA	CGCGACAGCG	CTCGCGCGCG	CGCGCGCGCG
951	CCTTTCCGTC	ARTTTGAAG	AACCGAACAA	GGCAACGGTT	TCTGCGGAGG
1001	CGCGGATTC	GGCGCTGATT	CCGSAAGTC	GGACGTTGT	CGSGAAACGG
1051	GATGTCGAAA	TGCGCTCTGA	AACCGAAAT	GTTTTCACGG	AAACCGTTTC
1101	GTCTGTGGGA	TACGGCGGTC	CGGTTTATGA	TGAAGCTGCC	GATATCCATA
1151	TTGAAGAGCC	TGCGCGCGCC	GATGCTTGGG	TGGTCGAACC	ACCGGAAGTG
1201	CGGAGGTAG	CGGTACCCGA	AATCGATATT	CTCGCGCGCG	CTCCCGTATC
1251	GGAAATCTAC	AACCGTACCT	ATGAGCGCGC	GGCAGGATTC	GAGCAGGGCG
1301	AACGAGCGCG	CATTGCCGAA	ACCGACCATC	TTGCGCTGA	TGTTTNTGAAT
1351	GGAGGTTGGC	AGGAGGAAC	CGCGCTATT	CGAGATGACG	GCACTGAGGG
1401	TGCGGCGAG	CGGTCAAGCG	GGCAATATCT	GTCGGAAACC	GAAGCGTTTG
1451	GGCATGACAG	TCAAGCGGTT	TGTCGTTTG	AAGATGTGCC	GCTCGAACCG
1501	CGGTCTGCC	GGGTATCGGA	TACGGAAGCG	GATGAAGGGG	CGTTCACATC
1551	GGAGAGGACC	GGTGGGTTAT	CGGAACACCT	CGCGCAACC	GACCTGCTTC
1601	TGCTTCGCT	GTTCATCCC	GAGGCGAGCG	AAACCGAAGA	AGAACTGTTG
1651	GRAAACAGCA	TCACCATCGA	AGAAAATTG	CGCGGATTC	AACTCAGGTT
1701	CAGGTTTGT	GATCTTATT	CGCGCGCGCT	GATTACGCGT	TATGAATCG
1751	AACCGCATGT	CGCGTGGCG	GGCAATTCCG	TTCTGATATT	GGAAAGAAC
1801	TTGCGCGGTT	CGCTCGCGCT	GGCTCCATC	CGCGTTGTC	AAACCATCCC
1851	CGGCAAAACC	TGCATGGGTT	TGGAACCTCC	GAACCCGAAA	CGCCAAATGA
1901	TACGCTGAG	CGAAATTTTC	AATTCGCGCG	AGTTTTCGGA	ATCCAAATCC
1951	AAGCTGACGC	TGCGGCTCGG	TCAGGACATT	ACCGGACAGC	CGGTGCTAAT
2001	CGACTTGGCG	AAAGCACCGC	ATTTCGTGCT	TGCGGCGAGC	ACCGGTTCCG
2051	GCAATCGGT	GGGTGTCAAC	GCGATGATTC	TGTCTATGCT	TTTCAAGGCC
2101	GCGCGGAAG	ACGTGGGTAT	GATTATGATC	GATCCGAAAA	TGCTGGAATT
2151	GAGCATTTAC	GAAGGCATCA	CGCACTGCT	CGCCCGTCT	GTTACCGGTA
2201	TGAAGCTGGC	GGCAACGCG	CTGAACCTGT	GTGTTAAGCA	AATGGAAAAA
2251	CGCTACCGCC	TGATGAGCTT	TATGGCGGTG	CGCAATCTTG	CGGGCTTCAA
2301	CCAAAAAATC	CGCGAAGCG	CAGCAAGGGG	AGAAAAAATC	GGCAATCCGT
2351	TCAAGCTCAC	CGCCGACGAT	CCGCAACCTT	TGGAAAAAAT	CGCGTTTATC
2401	GCTGCTGTGG	TGATGAGCTT	TGCGGATTTG	ATGATGACGG	CAGGCAAGAA
2451	AATCGAAGCA	CTGATGACCC	GCTCGCGCA	AAAAGCCGCG	CGCGCGAGCA
2501	TCCGACTTAT	CGTGGCGACA	CAGACGATGT	CGTCCGATGT	CATCCGAGGT
2551	CTGATTAAAG	CGAACATCCC	GACGCGATC	CGCTTCCAG	TGTCGACAA
2601	AATCGACAGC	CGCACGATTC	TGCAACCAAT	GGCGCGCGAA	ACCTGCTCTG
2651	GTACGGCGSA	TATGCTGTTT	CTCGCGCGCG	GTACTGCTTA	TCCGCGCGCG
2701	GTTCAAGGCG	CGTTTGCTTC	GGATGAAGAG	GTGCAACCGT	TGCTGCAATA
2751	TCTGAAGCAG	TTTGCGGAGC	CGACTATGT	TGACCATATT	TTGAGCGGGC
2801	GCGGACGAGA	AGAGCTGCC	GGCATCGGGC	GCAGCGCGGA	CGCGGAAACC
2851	GATCCGATGT	ACGACGAGGC	CGTATCGGTT	GTCTGTAATA	CGCGCAAGCC
2901	CAGCATTTGC	GGGCTACAGC	CGCGCTTGGC	CATCGGCTAC	AACCGCGCGC
2951	CGGCTGTGAT	TGACCAAAAT	GAGCGGGAAG	GCATTTGTTG	CGCACCGGAA
3001	CACAAAGGCA	ACCGTACGAT	TCTCGTCCCC	TTGGACAATG	CTTGA

This corresponds to the amino acid sequence <SEQ ID 496; ORF58ng-1>:

1	MEWIVLIVIV	LLALAGLFFV	RAQSEREWMR	EVSWAQEKKG	EKQAELEPIK
51	DGMPDFEFS	LMLFHAVKTA	VYWLFGVVR	FCRNYLAHES	EPDRPVPPAS
101	ANRADVPAS	DGYSDSNGMT	EEAEETAEEA	EEEAADTED	IATAVIDNRR
151	IFPDRSLAEG	LMQSEKSTSP	VRPVFKEITL	EEATRALSSA	ALRETKKRYI
201	DAFEKNGTRV	PKVRVSDTFM	EGLQIIGLDD	EVQRTYSYR	FDADKEAFSE
251	SADVGEPFVF	EPKEPBAFSA	VKAENRNAP	FRHMGQKRG	QAEAKSPDVS
301	QCGVSDGTA	VEDARRRVVS	NKEPKNKAT	SEARISRLI	EESEPVVCKR
351	DVEMPESETEN	VPTETVSVSG	YGVVVDVDA	DHIEEPFAF	DAMVVEPEVF
401	EVAVPEIDIL	LPFPVSEIY	NRTYEPAGF	VQAKRSRIAE	TDLHAADVNL
451	CGWQEEATAI	ADDGSEGNAE	RSQGYLSET	EAFGHDSDAV	CPFEDVPSE
501	PSCRVSDTEA	DEGAFQSEET	GAVSEHLPTT	DLLLPPLFNP	RATQTEEELL
551	ENSITIEEKL	AEFKVKVKV	DSYSGFVITR	YEIEPDVGV	GNSVLNLEKD
601	LARSLGVASI	RVVETIPGKT	CMGLELPNPK	QMILRESIF	NSPEFAESKS
651	KLTLALGQDI	TQGPVVTDLG	KAPHLVAGT	TGSGKSGVGN	AMILSMLFKA
701	AFEDVRMIMI	DPKMLELSIY	EGITHLLAPV	VTDMKLAANA	LNWCVNEMEK
751	RYRLMSFMVG	RNLAFGNQKI	AEAAARGEKI	GNFSLTPDD	PEPLEKLEFI
801	VVVVDEFADL	MTAGKQKIEE	LIALRLAQKAR	AGIHLILAT	QRPSVDVITG
851	LKANIIPTRI	AFQVSSKIDS	RTILDQMGAE	NLLGGQDMFL	LPPTATYVQR
901	VHGAFAFSD	VEHVVEYLKQ	FGEVDYVDDI	LSGGGSEELE	GIGRSGDGET
951	DEMYDEAVSV	VLRTRKASIS	GWQRALRIGY	NRAARLIDQM	EABGVTSAPB
1001	HNGNKTILVP	LQNA*			

ORF58ng-1 and ORF58-1 show 97.2% identity in 1014 aa overlap:

		10	20	30	40	50	60
	orf58-1.pep	MFWIVLIVILLALAGLFFVRAQSEREMREVSAWQEKQAEPEIKDKGMPDFPELA					
5	orf58ng-1	MFWIVLIVIVLLALAGLFFVRAQSEREMREVSAWQEKQAEPEIKDKGMPDFPEFS					
		10	20	30	40	50	60
	orf58-1.pep	LMLFHAVKTAVYWLFGVVRFCRNYLAHESEPDPRVPPASANRADVPPTASDGYSDSGNGT					
10	orf58ng-1	LMLFHAVKTAVYWLFGVVRFCRNYLAHESEPDPRVPPASANRADVPTASDGYSDSGNGT					
		70	80	90	100	110	120
	orf58-1.pep	EAEETEAEAAEEEAADTEDIATAVIDNRRIPFDRSIAEGLMPSESEISFVRVPKEITL					
15	orf58ng-1	EAEETEAEAAEEEAADTEDIATAVIDNRRIPFDRSIAEGLMQSESKTSFVRVPKEITL					
		130	140	150	160	170	180
	orf58-1.pep	EAEETEAEAAEEEAADTEDIATAVIDNRRIPFDRSIAEGLMPSESEISFVRVPKEITL					
	orf58ng-1	EAEETEAEAAEEEAADTEDIATAVIDNRRIPFDRSIAEGLMQSESKTSFVRVPKEITL					
		190	200	210	220	230	240
20	orf58-1.pep	EEATRALNSAALRETKKRYIDAFAEKNETAVPKVRVSDTPEMGLQIGLDDPVLQRTYSIM					
	orf58ng-1	EEATRALNSAALRETKKRYIDAFAEKNGTAVPKVRVSDTPEMGLQIGLDDPVLQRTYSIM					
		190	200	210	220	230	240
	orf58-1.pep	EEATRALNSAALRETKKRYIDAFAEKNETAVPKVRVSDTPEMGLQIGLDDPVLQRTYSIM					
	orf58ng-1	EEATRALNSAALRETKKRYIDAFAEKNGTAVPKVRVSDTPEMGLQIGLDDPVLQRTYSIM					
		250	260	270	280	290	300
25	orf58-1.pep	FDADKEAFSESADYGFEPYFEKQHPFSAFSAVKAENARNAPFRRHAGQKGQAEAKSPDVS					
	orf58ng-1	FDADKEAFSESADYGFEPYFEKQHPFSAFSAVKAENARNAPFRRHAGQKGQAEAKSPDVS					
		250	260	270	280	290	300
30	orf58-1.pep	GGQSVSDGTAVRDARRRVSVNLKEPNKATVSAEARISRLIPESQTVVGKRDVEMPSSETN					
35	orf58ng-1	GGQSVSDGTAVRDARRRVSVNLKEPNKATVSAEARISRLIPESQTVVGKRDVEMPSSETN					
		310	320	330	340	350	360
	orf58-1.pep	VFTETVSSVGYGGPVYDEADIIHIEEPAAPDAWVVEPPEVPKPMPTAIDIQPPPPVSEIY					
40	orf58ng-1	VFTETVSSVGYGGPVYDEADIIHIEEPAAPDAWVVEPPEVPEVAWPEIDILPPPPVSEIY					
		370	380	390	400	410	420
	orf58-1.pep	VFTETVSSVGYGGPVYDEADIIHIEEPAAPDAWVVEPPEVPEVAWPEIDILPPPPVSEIY					
	orf58ng-1	VFTETVSSVGYGGPVYDEADIIHIEEPAAPDAWVVEPPEVPEVAWPEIDILPPPPVSEIY					
		430	440	450	460	470	480
45	orf58-1.pep	NRTYEPFSGFEQVORSRIAETDHLADDVLNGGQWEEATAIADDGSEGAERSSGGYLSET					
	orf58ng-1	NRTYEPFAGFEQAQRSRIAETDHLAADVLNGGQWEEATAIADDGSEGAERSSGGYLSET					
		430	440	450	460	470	480
	orf58-1.pep	EAFGHDSQAVCPFENVPSESRPSCRVSDEADEGAFQSEETGAVSEHLPTDOLLPPFLNP					
50	orf58ng-1	EAFGHDSQAVCPFEDVPSESRPSCRVSDEADEGAFQSEETGAVSEHLPTDOLLPPFLNP					
		490	500	510	520	530	540
	orf58-1.pep	EATQTEELLENISITIEEKLAEFKVKVVDYSYSGPVITRYEIEPDVGVGRNSVLNLEKD					
	orf58ng-1	EATQTEELLENISITIEEKLAEFKVKVVDYSYSGPVITRYEIEPDVGVGRNSVLNLEKD					
		550	560	570	580	590	600
55	orf58-1.pep	EATQTEELLENISITIEEKLAEFKVKVVDYSYSGPVITRYEIEPDVGVGRNSVLNLEKD					
	orf58ng-1	EATQTEELLENISITIEEKLAEFKVKVVDYSYSGPVITRYEIEPDVGVGRNSVLNLEKD					
		610	620	630	640	650	660
60	orf58-1.pep	LARSLGVASIRVVETIPGKTCMGLELPNPKQMIRLSEIFNSPEFAESKSKLTLALGQDI					
65	orf58ng-1	LARSLGVASIRVVETIPGKTCMGLELPNPKQMIRLSEIFNSPEFAESKSKLTLALGQDI					
		610	620	630	640	650	660
	orf58-1.pep	TGQPVVTDLGKAPHLLVAGTTGSGKSVGVNAMILSMLFKAAPEDVVRMIDIPKMLEISY					
70	orf58ng-1	TGQPVVTDLGKAPHLLVAGTTGSGKSVGVNAMILSMLFKAAPEDVVRMIDIPKMLEISY					
		670	680	690	700	710	720

		730	740	750	760	770	780
	orf58-1.pep	EGIPHLAPVVTDMKLAANALNWCNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKI					
5	orf58ng-1	EGITHLLAPVVTDMKLAANALNWCNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKI					
		730	740	750	760	770	780
	orf58-1.pep	GNPFSLTDDPEPELEKLPFIVVVVDFADLMNTAGKKEELIARLAQKARAAGIHLILAT					
10	orf58ng-1	GNPFSLTDDPEPELEKLPFIVVVVDFADLMNTAGKKEELIARLAQKARAAGIHLILAT					
		790	800	810	820	830	840
	orf58-1.pep	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMAENLLGQGDMLFLPGTAYPQR					
15	orf58ng-1	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMAENLLGQGDMLFLPGTAYPQR					
		850	860	870	880	890	900
	orf58-1.pep	VHGFASDEEVRHVVEYLYKQFGEPEYVDDILSGGGSEELPGIGRSGDDETDMYDEAVSV					
20	orf58ng-1	VHGFASDEEVRHVVEYLYKQFGEPEYVDDILSGGGSEELPGIGRSGDDETDMYDEAVSV					
		910	920	930	940	950	960
	orf58-1.pep	VHGFASDEEVRHVVEYLYKQFGEPEYVDDILSGGGSEELPGIGRSGDDETDMYDEAVSV					
25	orf58ng-1	VHGFASDEEVRHVVEYLYKQFGEPEYVDDILSGGGSEELPGIGRSGDDETDMYDEAVSV					
		970	980	990	1000	1010	
	orf58-1.pep	VLKTRKASISGVQRALRIGYNRAARLIDQMAEGIVSAPEHNGNRTILVPLDNAX					
30	orf58ng-1	VLKTRKASISGVQRALRIGYNRAARLIDQMAEGIVSAPEHNGNRTILVPLDNAX					
		970	980	990	1000	1010	
	Furthermore, ORF58ng-1 shows significant homology to the <i>E. coli</i> protein FtsK:						
	sp P46809 FTSK_ECOLI CELL DIVISION PROTEIN FTSK >gi 1651412 gnl PID d1015290 (D1						
	division protein FtsK [Escherichia coli] >gi 1651418 gnl PID d1015296 (D90727) Cell						
	division protein FtsK [Escherichia coli] >gi 1787117 (AE000191) cell division						
35	protein FtsK [Escherichia coli] Length = 1329						
	Score = 576 bits (1469), Expect = e-163						
	Identities = 301/459 (65%), Positives = 353/459 (76%), Gaps = 5/459 (1%)						
	Query: 556 IEKLAEFKVKVVDVSYSGPVITRYEIEPDVGVGRNSVLNLEKDLARSLGVASIRVYET 615						
40	Sbjct: 868 VEARLADFRKADVVNYSFGVITRFELNLAFGVKAARISNLSRLARSLSTAVRVVVE 927						
	Query: 616 IPGKTGMLPKNPKRMQIRLSEIFNSPEAESKLTALGQDITGQPVVTDLGAAPHL 675						
45	Sbjct: 928 IPGKFGVLELPNKRQTVYLRVLEVLNAKFRDNPSPLTVVLGKDIAEPVADLAKMPEH 987						
	Query: 676 LVAGTTSGGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIYEGITHLLAPVVTDMK 735						
50	Sbjct: 988 LVAGTTSGGKSVGVNAMILSMLYKAQPEDVRFIMIDPKMLELSIYEGITHLLAPVVTDMK 1047						
	Query: 736 LAANALNWCNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKIGNPFSLTDDPEP-- 793						
55	Sbjct: 1048 DAANALNWCNEMERRYKLSALGVNRNLAGYNEKIAEADRMRPIDPDYWKFGDSMDAQH 1107						
	Query: 794 --LEKLPFIVVVVDFADLMNTAGKKEELIARLAQKARAAGIHLILATQRPSVDVITGL 851						
60	Sbjct: 1108 PVLKKEFYIVLVDFADLMNTAGKKEELIARLAQKARAAGIHLVLAATQRPSVDVITGL 1167						
	Query: 852 IKANIPTRIAFQVSSKIDSRTILDQMAENLLGQGDMLFLPGTAYPQRVHGAFAASDEE 911						
65	Sbjct: 1168 IKANIPTRIAFVSSKIDSRTILDQGAESLLGMGDMILYSGPNSTLPVRVHGAFAVRDQEV 1227						
	Query: 912 HRVVEYLYKQFGEPEYVDDILSGGGSEELPGIGRSGDDETDMYDEAVSVVLKTRKASISG 971						
	Sbjct: 1228 HAVVDQWKARGRPQYVDGITSDESEGGAG-GFDGAELDPLFDQAVQFVTEKKRKASISG 1286						
	Query: 972 VQRLRIGYNRAARLIDQMAEGIVSAPEHNGNRTILV 1010						
	Sbjct: 1287 VQRQFRIGYNRAARLIDQMAEQAGVSEHGNGNREVLP 1325						

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 59

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 497>:

```

5      1  ATGATTIATC  AAAGAAACCT  CATCAAGAA  CTCTCTTTTA  CCGCGTCGG
    51  CATTTTCGTC  GTCCCTCTGG  CGGTATTGGT  CTCACGACG  GCAATCAAC
   101  TGCTCGGCCG  TGCCGCCGAC  GGGC..GTGA  TCGCCATCGA  TGCCGTGTTG
   151  GCATTGGTCG  GCTTCTGGGT  C.....

10     901  .....A  TTGCCATCGG  TTGTTTTTA  ATTTACCAAA  ACGGGCTGAC
    951  CCTGCTTTTT  GAAGCCGTGG  AAGACGGCAA  AATCATTTTT  TGGCTCGGAC
   1001  TGCTGCCTAT  GCACATTATC  ATGTTTGTCC  TTGCACTCAT  CCGTTTGGCG
   1051  GTCCCGCAGTA  TGCCACAGCA  GCCCTTCTGG  CAGGCGGTTG  GCAAAAGTCT
   1101  GACATTGAAA  GCGCGAAAAA  GA

```

15 This corresponds to the amino acid sequence <SEQ ID 498; ORF101>:

```

      1  MIYQRNLKE  LSFTAVGIFV  VLLAVLVSTQ  AINLLGRAAD  GXVIAIDAVL
    51  ALVGFVW... ..

   301  ...IAIGLFL  IYONGLTLLF  EAVEDGKIHF  WLGLLPMHII  MFVLAILILLR
   351  VRMSPQPFW  OAVGKSLTLK  GKK*

```

Further work revealed the complete nucleotide sequence <SEQ ID 499>:

```

      1  ATGATTTIATC  AAAGAAACCT  CATCAAGAA  CTCTCTTTTA  CCGCGTCGG
    51  CATTTTCGTC  GTCCCTCTGG  CGGTATTGGT  CTCACGACG  GCAATCAAC
   101  TGCTCGGCCG  TGCCGCCGAC  GGGC..GTGA  TCGCCATCGA  TGCCGTGTTG
   151  TGCTCGGCCG  TGCCGCCGAC  GGGC..GTGA  TCGCCATCGA  TGCCGTGTTG
   201  CGCATTTIATC  AGTACGTTGA  CCGTGTGTAC  CGCGTACTGG  CCGCACAGCG
   251  AAATGTCGGT  CTGGCTATCC  TCGCGATTGG  CATTTGAAAC  ATGATACGCG
   301  CCGGTGATGC  AGTTTGCCGT  GCGCGTTTGC  GTTTTGGTTG  CGGTATCGCA
   351  GCTTTGGGTG  ATACCGTGGG  CAGAGCTACG  CAGCGCGGAA  TACGTGAAAC
   401  TCCTGAAGCA  GAAGCAGGAA  TTGTCTTTGG  TGGAGGCGAG  CAGGTTCAAC
   451  AGTTTGGGCA  AGCGCAACGG  CAGGGTTTAT  TTTGTGGAAC  CCTTCGATAC
   501  CGAATCCGGC  ATCATGAAAA  ACCTGTTCTC  GCGCGAACAG  GACAAAAACG
   551  GCGGCGACAA  CATCATCTTC  GCCAAGAAG  GTAACTTCTC  GCTGAACGAC
   601  ACAAACGCA  CGCTCGAATT  GCGCCACGGC  TACCGTTTAC  GCGGCAAGCC
   651  CGGACGCGCC  GACTACAATC  AGGTTTCCCT  CCAAAACACT  AACCTGATTA
   701  TCAGCACCAC  GCCCAAACTC  ATCGACCCCG  TTTCCACACG  CGGTACCAAT
   751  CCGACGCGCC  AACTGATTGG  CAGCAGCAAC  CCGCAACATC  AGGCGGAATT
   801  GATGTGCGCG  ATCTCGCTGA  CGGTGACGCT  CCTCTACTCT  TGCGTCTGTT
   851  CGGTGCGGCT  TTCTATTTC  AACCCGCGCA  GCGGACATAC  CTACAATATC
   901  TTGATTGCCA  TCGGTTTGT  TTTAATTTC  CAAACCGGCG  TGACCTCGGT
   951  TTTTGAGGCC  GTGGAGACG  GCAATATCA  TTTTGTGCTC  GGACTCTGCT
  1001  CTATGCACAT  TATCATGTTT  CGCGTTGCAC  TCATCCTGTT  CCGCGTCCCG
  1051  AGTATGCCCA  GCCAGCCCTT  CTGGCAGCGC  GTTGGCAAAA  GTCTGCACAT
  1101  GAAAGCGGGA  AATGA

```

45 This corresponds to the amino acid sequence <SEQ ID 500; ORF101-1>:

```

      1  MIYQRNLKE  LSFTAVGIFV  VLLAVLVSTQ  AINLLGRAAD  GRVAIDAVLA
    51  LVGFVWVGMT  PLLLVLTAFI  STLTVLTRYW  RDSEMSVWLS  CGLALKQWIR
   101  PVMQFAVPEA  VLVAVMQLWV  IPWAEIERSR  YAEILKQKE  LSLVEAGEFN
   151  SLGKRNRVY  FVETFDTEG  IMKNLFLREQ  DRNGDNIIF  AKEGNFSLND
   201  NKRTLRLRHG  YRYSGPFGRA  DYNQVSFKL  NLIISTTPKL  IDPVSHRRTI
   251  PTAQLIGSSN  PQHQAEIMWR  ISLTVSVLLL  CLLAVPLSYF  NPSRGHTYNI
   301  LIAIGLFLIY  ONGLTLLFEA  VEDGKIFHWL  GLLPMHIME  AVALILLRVR
   351  SMPSPQFWQA  VGKSLTLKGG  K*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF101 shows 91.2% identity over a 57aa overlap and 95.7% identity over a 69aa overlap with an ORF (ORF101a) from strain A of *N. meningitidis*:

		10	20	30	40	50
5	orf101.pep	MIYQRNLKELSF	TAVGIFVVL	LAVLVSTQAINLL	GRAADGXVIAID	DAVLVGFVWX
	orf101a	MIYQRNLKELSF	TAVGIFVVL	LAVLVSTQAINLL	GXAADXRX-AID	DAVLVGFVWXXM
		10	20	30	40	50
				//		
10	orf101.pep	IAIGLFLIYQ	NGLTLLFE	AVEDEGKIHFWL	GL
	orf101a	LTVS	VLLCLLAVPLSYFN	PRSGHTYNIL	KXAI	GLFLIYQNGLTLLFE
		280	290	300	310	320
15	orf101.pep
	orf101a
		120	130	140	150	
	orf101.pep	LPMHII	MFVLA	ILLVR	SMSPQPF	WQAVGKSLTLKGGKX
	orf101a	LPMHII	MFVLA	ILLVR	SMSPQPF	WQAVGKSLTLKGGKX
		340	350	360	370	

The complete length ORF101a nucleotide sequence <SEQ ID 501> is:

	1	ATGATTATC	AAAGAAACCT	CATCAAGAA	CTCTCTTTTA	CCGCGCGTCGG
	51	CATTTTGGT	GTCCTCTGG	CGGTATTGGT	CTCCACGCAG	GCAATCAACC
	101	TGCTCGGCCN	TGCGCCGAC	NGCGTNTCG	CCATCGATGC	CGTGTGGCA
25	151	TTGTCGGCT	TCTGGTCNN	NNGNATGAC	CGCTTTTGC	TNGTGTGAC
	201	CGCATTTATC	AGTACGTGA	CGTGTTGAC	CGCTACTGG	CGNACACGG
	251	AAATGTCGGT	CTGGNTATCC	TGCGGATTGG	CATTGAACA	ATGGATACCG
	301	CCGSGTATGC	AGTTTGGCGT	GCGGTTTGGC	GTTTGTGTTG	CCGTGATGCA
30	351	GCTTTGGTG	ATACCGTGG	CAGACGACG	CACGCGCAA	TACGCTGAAA
	401	TCTCTAGCA	GAACGAGAA	TGCTCTTGG	TGGAAGCAG	CGGGTACGCA
	451	AGTTTGGCA	AGCGCAACGG	CAGGTTTAT	TTTGTGCAA	CCTTGATATC
	501	CGAATCCGG	ATCATGAAA	ACCTGTTCT	GCGGACACG	GACAAAAACG
	551	GCGGCGCAA	CATCATCTTC	NCCAAAGAAA	GTAACCTCTC	GCTGACACG
	601	AACAAACGA	CGCTCGAAT	GCGCCACGG	TACCGTTACA	GCGGACGCG
35	651	CGGACGCGC	GACTACAATC	AGGTTTCTT	CNNAARCTC	AACCTGATTA
	701	TCAGACCCAC	GCCCAACTC	ATCGACCGG	TTTCCACCG	CGTACNATN
	751	CNACNCGCC	AACTGATTGG	CAGCAGCAAC	CGCACAATC	ANGCGGAATT
	801	GATGTGGCG	ATCTCGCTGA	CGTCAGCGT	CCTCTACTC	TGCTGCTTG
40	851	CGGTGCGCT	TTCTATTTC	AACCCGCGCA	GCGGACATAC	CTACAATATC
	901	TTGATGCCA	TGGTTTGTG	TTTAATTTC	CAAAACGGG	TGACCGTGCT
	951	TTTTGAAGCC	GTGGAAGACG	GCAAAATCCA	TTTTTGCTC	GGACTGCTGC
	1001	CTATGCACAT	CATCATGTTC	GTCATCGCA	TGCTACTTCT	GCGGTCGCG
	1051	AGCATGCCA	GCGAGCCCT	CTGCGAGCG	GTGGCAAAA	GTCTGACATT
	1101	GAAGGCGGA	AAATGA			

45 This encodes a protein having amino acid sequence <SEQ ID 502>:

	1	MIYQRNLKEL	LSFTAVGIFV	VLLAVLVSTQ	AINLLGXAAD	XRXAIDAVLA
	51	LVGFVWXXMT	PLLLVLTAFI	STLTVLTRYK	RDSMSVWKS	CGLALKQWIR
	101	PVMQFAVPEA	VLLAVVQLWV	IPWALRLSRE	YAILKQKQE	LSLVEAGGFN
	151	SLGRNGRVY	FVETFTESG	IMKNLFLREQ	DKNGGDNIIF	XKESNFSIAD
50	201	NKRTLRLHC	YRYSPTPGRA	DYNQVSFKL	NLIISTTPKL	IDPVSHRRTX
	251	PTAOLIGSSN	POHKAELMR	ISLTVSVLL	CLLAVPLSYF	NPRSGHTYNI
	301	LXAI	GLFLIYQ	NGLTLLFE	AVEDEGKIHFWL	GLLPMHII
	351	SMSPQPF	WQAVGKSL	TLKGG	K*	

ORF101a and ORF101-1 show 95.4% identity in 371 aa overlap:

55	orf101a.pep	MIYQRNLKELSF	TAVGIFVVL	LAVLVSTQAINLL	GXAADXRXAID	DAVLVGFVWXXMT	60
	orf101-1	MIYQRNLKELSF	TAVGIFVVL	LAVLVSTQAINLL	GRAADGRVAID	DAVLVGFVWIGMT	60
	orf101a.pep	PLLLVLTAFI	STLTVLTRYWRD	SEMSVWKS	CGLALKQWIRPVMQFAV	PAVLVAVMQLWV	120
60	orf101-1	PLLLVLTAFI	STLTVLTRYWRD	SEMSVWKS	CGLALKQWIRPVMQFAV	PAVLVAVMQLWV	120

5	orf101a.pep	IPWAELRSREYAEILKQKQELSLVEAGGNSLGKRNQGVYFVETFDTESGIMKNLFREQ	180
	orf101-1	IPWAELRSREYAEILKQKQELSLVEAGEINSLGKRNQGVYFVETFDTESGIMKNLFREQ	180
10	orf101a.pep	DKNGGDNIIFXKESNFSLNDNKRTLELRHGYRYSYSGTPGRADYNQVSFXKLNLIISTTPKL	240
	orf101-1	DKNGGDNIIFAKEGNFSLNDNKRTLELRHGYRYSYSGTPGRADYNQVSFQKLNLIISTTPKL	240
15	orf101a.pep	IDPVSHRRITPTAQLIGSSNPQHKAELMWRISLTVSVLLCLLAVPLSYFNPRSGHTYNI	300
	orf101-1	IDPVSHRRITPTAQLIGSSNPQHKAELMWRISLTVSVLLCLLAVPLSYFNPRSGHTYNI	300
20	orf101a.pep	LKAIGFLIYQNGLTLLFEAVEDGKIHFVGLLPMHIIMFVIAIVLLRVRSMPSQPFWQA	360
	orf101-1	LKAIGFLIYQNGLTLLFEAVEDGKIHFVGLLPMHIIMFAVAILLLRVRSMPSQPFWQA	360
	orf101a.pep	VGKSLTLKGGK 371	
	orf101-1	VGKSLTLKGGK 371	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF101 shows 96.5 % identity in 57aa overlap at the N-terminal domain and 95.1% identity in 61aa overlap at the C-terminal domain, respectively, with a predicted ORF (ORF101ng) from *N.*

gonorrhoeae:

30	orf101.pep	MIYQRNLIKESFTAVGIFVVLAVLVSTQAINLLGRAADGXVIAIDAVLAVGVFW	57
	orf101ng	MIYQRNLIKESFTAVGIFVVLAVLVSTQAINLLGRAADGRV-AIDAVLAVGVFWIGM	59
35	orf101.pep	IAIGFLIYQNGLTLLFEAVEDGKIHFVGL 333	
	orf101ng	SLTVSVLLCLLAVPLSYFNPRSGHTYNI LIAIGFLIYQNGLTLLFEAVEDGKIHFVGL 331	
	orf101.pep	LLPMHIIMFVLAILLRVRSMPSQPFWQAVGKSLTLKGGK 373	
	orf101ng	LLPMHIIMFVIAIVLLRVRSMPSQPFWQAVG 362	

The ORF101ng nucleotide sequence <SEQ ID 503> is predicted to encode a protein having partial amino acid sequence <SEQ ID 504>:

45	1	MIYQRNLIKE	LSFTAVGIFV	VLLAVLVSTQ	AINLLGRAAD	GRVAIDAVLA
	51	LVGFVIGMT	PLLLVLTAFI	STLTVLTRYW	RDSEMSVWLS	CGLAKQWIR
	101	FVMQFAVFFA	LLAVMQLMW	IPWAELRSRE	YAEILKQKQE	LSLVEAGEFN
	151	NLGRNQRVY	FVETFDTESG	IMKNLFLREQ	DKNGGDNIIF	AKEGNFSLDK
	201	NKRTLELRHG	YRYSYSGTPGRA	DYNQVSFQKL	NLIISTTPKL	IDPVSHRRIT
	251	STAQLIGSSN	PQHKAELMWR	ISLTVSVLLCL	LAVPLSYFN	NPRSGHTYNI
	301	LIAIGFLIY	QNGLTLLFEA	VEDGKIHFVGL	GLLPMHIIMF	VIAIVLLRVR
	351	SMPSQPFWQA	VG...			

Further work revealed the complete nucleotide sequence <SEQ ID 505>:

50	1	ATGATTATC	AAAGAAACCT	CATCAAGAA	CTCTCTTTTA	CCGCGCTGG
	51	CATTTGCTC	GTCTCTTGG	CGGTGTTGGT	GTCCACGAG	GGATCAACC
	101	TGCTTGGCG	CGCAGCTCAC	GGGCTGTGCT	CCATGATGTC	CGTGTGGCC
	151	TTACTCGGCT	TCTGGGTGAT	CGGTATGACC	CCGCTTTTGC	TGGTGTGAC
55	201	CGCATTCATC	AGCAGCTGTA	CCGTATTGAC	CCGCTACTCG	CGCAGAGGS
	251	AAATGTCGGT	CTGCTATACC	TGCGGATTGG	CGTTGAACA	GTGATACGC
	301	CCGCTCATGC	AGTTTGGCGT	CGCGTTTGGC	ATCCTGATTG	CCGTATGCA
	351	GCTTTGGGTG	ATACCGTGGG	CAGAGCTGCG	CAGCCGCGAA	TATGCGGAA
60	401	TTTGAAGCA	GAAGCAGGAA	TTGCTTTTGG	TGGAAGCCGG	CGAGTTCAT
	451	AACTTGGGCA	AGCGCAACGG	CagggtttaaT	TtctcgaaaA	CCTTTGACAC
	501	GgaatcgGC	ATCATGAAA	ACCTGTtact	GcGGAACAG	GACAAAAGC
	551	gcgcgacacA	CATCATCTTC	GcCaagAag	gtaactTctc	gctgaaggacG

```

601 AACAAAcgca cgcctgaATT GCGCCACGGC TACCGTTACA GCGGcagcgC
651 CGGacGCGC gactCAATC AGGTTtccT GCAPAACTC aactcgtTt
701 TCAGCACCAC GCCCAaactT ATGaccCGG TTTCCACCG CCGGACCAAT
751 tcgacCGCCC AactGATTGG CAGCAGCAAT CGCAGACAT AGGCAGAAAT
5 801 GATGTGGCGC ATCTCGCTGA CCGTCAGCGT CCTCTGCTC TGCCTACTGC
851 CCGTGGCGCT TCTCTATTTC AACCAGCGCA GCGGACATAC CTACAATATC
901 TTGATTGCCA TCGGTTTGT TTTAATTAC CAAACGGGC TGACCTGTGT
951 TTTTGAAGCC GTGGAAGACG GCAAAATCCA TTTTGGCTC GGACTGCTGC
10 1001 CTATGCACAT CATCATGTTC GTACATCGCA TCGTACTTCT GCGCGTCCGC
1051 AGTATGCCCA GCGACCCCTT CTGGCAGCGG GTTGGCAAAA GTCTGCACAT
1101 GAAAGcgGGA AAATGA

```

This corresponds to the amino acid sequence <SEQ ID 506; ORF101ng-1>:

```

1  MIYQRNLLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GRVAIDAVLA
51 LVGFVWIGMT PLLLVLTAFT STLTVLTRYW RDSEMSVWLS CGLALKQWIR
15 101 PVMQFAVPFA ILIAVMQLWV IPWAELRSRE YAEILKQKQE LSLVEAGEFN
151 NLGKRNGRVY FVETFDTESG IMKNLFREQ DKNGGDNIIF AKEGNFSLKD
201 NKRTLELRHG YRYSGPGRA DYNQVSFQKL NLIISTTPKL IDPVSHRRTI
251 STAQIGSSN POQAELMWR ISLTVSVLL CLLAVPLSYF NPRSGHTYNI
301 LIAIGFLIY QNGLTLL FEA VEDGKIFWL GLLPMIIMF VIAIVLLRVR
20 351 SMPSQPFWQA VGKSLTKG K*

```

ORF101ng-1 and ORF101-1 show 97.6% identity in 371 aa overlap:

```

                10      20      30      40      50      60
25 orf101-1.pep  MIYQRNLLIKELSFTAVGIFVVLLAVLVSTQAINLLGRAADGRVAIDAVLALVGFVWIGMT
orf101ng-1      MIYQRNLLIKELSFTAVGIFVVLLAVLVSTQAINLLGRAADGRVAIDAVLALVGFVWIGMT
                10      20      30      40      50      60
30 orf101-1.pep  PLLLVLTAFTSTLTVLTRYWRDSEMSVWLSCGLALKQWIRPVMQFAVPFAVLAVVMQLWV
orf101ng-1      PLLLVLTAFTSTLTVLTRYWRDSEMSVWLSCGLALKQWIRPVMQFAVPFAVLAVVMQLWV
                70      80      90      100     110     120
35 orf101-1.pep  IPWAELRSREYAEILKQKQELSLVEAGEFNSLAGKRNGRVYFETFDTESGIMKNLFREQ
orf101ng-1      IPWAELRSREYAEILKQKQELSLVEAGEFNSLAGKRNGRVYFETFDTESGIMKNLFREQ
                130     140     150     160     170     180
40 orf101-1.pep  DKNGGDNIIFAKEGNFSLNDNKRTLELRHGYRYSGTPGRADYNQVSFQKLNLIISTTPKL
orf101ng-1      DKNGGDNIIFAKEGNFSLNDNKRTLELRHGYRYSGTPGRADYNQVSFQKLNLIISTTPKL
                190     200     210     220     230     240
45 orf101-1.pep  IDPVSHRRTIPTAQIGSSNPOQAELMWRISLTVSVLLCLLAVPLSYFNPRSGHTYNI
orf101ng-1      IDPVSHRRTIPTAQIGSSNPOQAELMWRISLTVSVLLCLLAVPLSYFNPRSGHTYNI
                250     260     270     280     290     300
50 orf101-1.pep  LIAIGFLIYQNGLTLLFEAVEDGKIFWLGLLPMIIMFAVAILLLRVRSMPSQPFWQA
orf101ng-1      LIAIGFLIYQNGLTLLFEAVEDGKIFWLGLLPMIIMFAVAILLLRVRSMPSQPFWQA
                310     320     330     340     350     360
55 orf101-1.pep  VGKSLTKGGKX
orf101ng-1      VGKSLTKGGKX
                370
60 orf101-1.pep  VGKSLTKGGKX
orf101ng-1      VGKSLTKGGKX
                370

```

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is

predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 60

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 507>:

```

5      1  ..GGTGGTGGTT TTATCAATGC TTCCTGTGCC ACTTTGACGA CAGCCAAACC
      51  GCAATATCAA GCGAGAGACC TTACGCGTTT TAAGATAAGC CAAGGCAATG
     101  TTGTAATCCG CGGACACGGT TTGGATGCAC GTATACCGA TTACACACGT
     151  ATTTCTCAGTT ATCATTCCAA AATCGATGCA CCGTATGGG GACAAGATGT
     201  TCGTGTGCTG GCGGGACAAA ACGATGTGGC CGCAACAGGT GATGCACATT
    10  251  CGCCTATTCT CAATAATGCT GCTGCCAATA CGTCAACAAA TACAGCCAAC
     301  AACGGCACAC ATATCCCTTT ATTTGCGATT GATACAGGCA AATTAGGAGG
     351  TAT.GTATGC CAACAAATC ACCTTGATCA GTACGGTGA GCAAGCAGGC
     401  ATTCGTAA
  
```

This corresponds to the amino acid sequence <SEQ ID 508; ORF113>:

```

15      1  ..GGGFINASCA TLTTAKPOYO AGDLSAFKIR QGNVVIAGHG LDARDTDYTR
      51  ILSYHSKIDA FVWGQDVRVV AGQNDVAATG DAHSPILNNA AANTSNNAN
     101  NGTHIPLFAI DTGKLGKXVC QQNHLDOYGR ASRHS*
  
```

Computer analysis of this amino acid sequence gave the following results:

Homology with with pspA putative secreted protein of *N.meningitidis* (accession AF030941)

20 ORF and pspA show 44% aa identity in 179aa overlap:

```

      orf113  GGGFINASCATLTTAKPOYQAGDLSAFKIROGNVVIAGHGLDARDTDYTRILSYHSKIDA 60
      GGG INA+ TLT+ P G+L+ F+ G VVI G GLD D DYTRILS ++I+A
      pspa    GGGILNAAVSILTSVLPVNLNGLTGFVDSGKGVVIGGKGLDTSADYTRILSRAAEINA 256

    25      orf113  FVWGQDVRVVAGQNDVAATGDAHSPILXXXXXXXXXXXXXGTHIPLFAIDTGLKGMYYA 120
      VWG+DV+VV+G+N + G + P AIDT LGMYA
      pspa    GVWGKDVVVSCKNKLDFD-----SLAKTASAPSSSDSVTPTVAIDTATLGMYA 307

    30      orf113  NKITLISTVEQAGIRNQGFASAGNVAVNAEGKLVNTGMIAATGENHAVSLHARNVHN 179
      +KITLIST A IRN+G+ FA+ G V ++A+GKL N+G I A +++ A+ V N
      pspa    DKITLISTDNGAVIRNKRIFAATGGVTLSDGKLSNSGSDIAA----ETTISAQTVDN 362
  
```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF113 shows 86.5% identity in 52aa overlap at the N- terminal part and 94.1% identity in 17aa

35 overlap at the C-terminal part with a predicted ORF (ORF113ng) from *N. gonorrhoeae*:

```

      orf113  GGGFINASCATLTTAKPOYQAGDLSAFKIR 30
      orf113ng SHPSQLNGYIEVGRRAEVVIANPAGIAVNGGGFINASRATLTGQPOYQAGDFSGFKIR 224

    40      orf113  QGNVVIAGHGLDARDTDYTRILSYHSKIDAFVWGQDVRVVAGQNDVAATGDAHSPILNNA 90
      orf113ng QGNAVIAGHGLDARDTDFTRILVCCQNHLDOYGRTSRHS 263

    45      orf113  IDTGKLGKXVCQQNHLDOYGRASRHS 135
      orf113ng DFGFKIROGNAVIAGHGLDARDTDFTRILVCCQNHLDOYGRFSRHS 263
  
```

The complete length ORF113ng nucleotide sequence <SEQ ID 509> is predicted to encode a protein having amino acid sequence <SEQ ID 510>:

```

50      1  MNKTLRYVIF NRKRGAVVAV AETTKREGKS CADSGSGSVY VKSVSFIFTH
      51  SKAPCFSA LG FSLCLALGTV NIAFADGIIT DKAAPKTQQA TILQTNGPIF
  
```

```

101 QVNIQTPTSA GVSVNQYACF DVGNRGAILN NRSRSTOTOI GGMWQGNPWL
151 TRGEARVVVN QINSSHPOL NQYIEVGGRI AEUVIANFAG TAVNGGGFTN
201 ASRATLLTGG PQYQAGDFSG FKIRQGNNAVI AGHGLDARDT DFTRILVCQQ
251 NHLDOYGRTS RHS*

```

- 5 Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 61

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 511>:

```

10 1 ..TCAACGGGAC ATAGCGAACA AAATTACACT TTGCCGCGAG AAATCACAGC
51 CAACATTTCAT CTGGGTTTCAT TTGCCTATGA ATCGCATCGC AAAGCATTAA
101 GCCATCATGC GCCACGCCAA GGCCTAGTGT TGCCGCAAAAG CAACGGTATT
151 TCGCTACCCCT ATACGTCCAA TTCTTTTACC CCATTACCCA GCAGCAGCTT
201 ATACATTATC AATCCTGTCA ATAAAGGCTA TCTTGTGAA ACCGATCCAC
251 GCTTTGCGAA CTACCGTCAA TGTTTGGTGA GTGACTATAT GCTGGACAGC
15 301 CTCAACACTAG ACCCAACAAA TTTACATAAA CTTTGGGTGT ATGGTTTATA
351 CAGCAACACT TTAATCAATG AACAAATCGC AGAGCTAGCA GGCATCGCT
401 GTTGAAGCGG TTATCAAAAC GACGAAGACG AATTTAAAGC CTTAATGGAT
451 AATGCGCGGA CTGCGGCACG TTcGATGAAT CTCAGCGTTG GCATTGCAAT
501 AAGTGCAGCG CAAGTAGCCG AACTGACCAG CGATATTGTT TGGTTGGTAC
20 551 AAAAAGAAGT TAAGCTTCTT GATGGCGGCA CACAACCGT ATTGGTGCCA
601 CAGGTTTATG TACGCGTTAA AAATGCGCAG ATAGACGGTA AAGGTGCATT
651 GTTGTGAGCG AGCAATACAC AAATCAATGT TTCAGGCGAG CTGAAAACCT
701 CAGGCACGAT TGCAGGCGCG AATGCGCTTA TTATCAATAC CGATACGCTA
751 GACAAATATC GTGGGCGTAT TCATGCGCAA AAATCAGCGC TTACGGCCAC
25 801 ACAAGACATC AATAATATTG GCGGCATGCT TTCTGCGGAA CAGACATTAT
851 TGCTCAACGC AGGCAACAAAC ATCAACAGCC AAAGCACCC CGCCAGCAGT
901 CAAAATACAC AAGCGACGAG CACCTACCTA GACCCGAATG CAGGTATTTA
951 TATCACAGGC AAGAAAAAAG GTGTTT..

```

This corresponds to the amino acid sequence <SEQ ID 512; ORF115>:

```

30 1 ..STGHSEQNYT LPREITRNIIS LGSFAYESHR KALSHHAFSQ GTELFQSNGI
51 SLPYTNSNFT PLPSSSLYII NPVNKGVLVE TDRFANYRQ WLGSYMLCS
101 LKLDENNLEK RLGDGYEYQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
151 NGATAARSMN LQEGIALSAE QVAQLTSDIV WLQKEVKLE DGGTQTVLVE
35 201 QVYVRVNGND IDGKGLLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
251 DNIGRIHAQ KSAVTATQDI NNIGWMLSAE QTLNLAGNWN INQSTTASS
301 QNTQGSSTYL DRMAGIYITG KEKGV..

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the pspA putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF115 and pspA protein show 50% aa identity in 325aa overlap:

```

40 Orf115: 1 STGHSEQNYTLPREITRNIISLGSFAYESHRKALSHHAFSQGTLELFQSNGISLPEYTSNFT 60
    STG+S Y E++ +I +G AY+ + + F + NGI +T
pspA: 778 STGYSRSPYEPAFEVS-SIRMGISAYKGAPQQAIDPGTVVPVVEANGIHPTFT----- 831

45 Orf115: 61 PLPSSSLYIINEPVNKGVLVETDRFANYRQWLGSYMLDSLKLDPNNLHKKRLGDGYEYQR 120
    LP-SSL+ I P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYEYQ+
pspA: 832 -LPNSSLFAIAPNNKGYLIETDEPATDYRKWLGSYMLAALQDPNNIHKRLGDGYEYQK 890

50 Orf115: 121 LINEQIAELTGHRRLDGYONDEEQFKALMDNGATAARSMNLVSGIALSAEQVAQLTSDIV 180
    L+NQIA+LFG+RRLDGY NDEEQFKALMDNG T A+ +L+ GIALSAEQVA+ITSDIV
pspA: 891 LVNEQIAKLTYRRLDGYTNDDEQFKALMDNGITAKKQLTGPGLIASAEQVARITSDIV 950

Orf115: 181 WLQKEVKLPDGGTQTIVLPQVYVRVKNIGDIDGKALLSGSNTQINVSGSLKN-SGTIAG 239
    WL + V LPDG TQTVL P+VYVR + D++G+ALLSGS I SG+++N G IAG
pspA: 951 WLENETVTLPDGTTQTVLKPVKYVVRARPKDMNGQALLSGSVVDIG-SGAENRGGLIAG 1009

```


Orf115: 240 RNALIINTDTLDNIGGRIHAQKSAVTATQDINNIGMLSAEQTLNAGXXXXXXX 299
 R ALI+ N + N+ G + + A DI N G + AE LLL A
 pspA: 1010 REALLIINAQIKNIQGDLDGKNIFAAAGSDITNTGS-TGAENALLKASNNIESRSETR 1068

5 Orf115: 300 XXXXXXXXXVLDRMAGIYITGKEG 324
 + R+AGIY+TG++ G
 pspA: 1069 NQNEQGSVRNIGRVAGIYLTGRQNG 1093

Homology with a predicted ORF from *N.gonorhoeae*

10 ORF115 shows 91.9% identity over a 334aa overlap with a predicted ORF (ORF115ng) from *N.gonorhoeae*:

orf115.pep	STGHSSEQNYTLPREITRNISLGSFAYESHK	31
orf115ng	NEQTGFGEKKVFSENGLHNYWRARRKGHDETGHEQNVTLPREETRDISLGSFAYESHK	71
orf115.pep	ALSHFAPSQGTLPQSN-----GISLPYTSNSTPLPSSSLYIINPVNKGYLVT	81
orf115ng	ALSRHAPSQGTLPQSNRDNIRTAKSNGISLPYTFNSFTPLPGSSLYIINPANKGYLVT	131
orf115.pep	DRPFANYRQWLGS DYMLDSLKLDPNNLHKRLGDSYEQRLINEQIAELTGHRRLDGYQND	141
orf115ng	DRPFANYRQWLGS DYMLGSLKLDPNNLHKRLGDSYEQRLINEQIAELTGHRRLDGYQND	191
orf115.pep	EEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWLVOKEVKLPDGGTQTVLVLP	201
orf115ng	EEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWLVOKEVKLPDGGTQTVLVLP	251
orf115.pep	VYVRVKNNGIDGKGALLSGSNQTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQK	261
orf115ng	VYVRVKNNGIDGKGALLSGSNQTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQK	311
orf115.pep	SAVTATQDINNIGMLSAEQTLNAGNNINSQSTTASSQNTQGSSTYLDRMAGIYITGK	321
orf115ng	SAVTATQDINNIGGILSAEQTLNAGNNINNSQSTAKSSQNAQGSSTYLDRMAGIYITGK	371
orf115.pep	EKG	325
orf115ng	EKGVLAAQAGKDINIIAGQISNSQDQGTQLQAGRDINDLTQTGKYQEIHFADNHTIR	431

An ORF115ng nucleotide sequence <SEQ ID 513> was predicted to encode a protein having amino acid sequence <SEQ ID 514>:

1	MLVQTEKDLG	HNEQTFGEKK	VFSENGKLN	YWFARRKGHD	ETGHEQNYT
51	LPEEITRDIS	LGSFAYESH	KALSRHAPSQ	GTELPQSNRD	NIRTAKSNGI
101	SLPYTPNSFT	PLFGSSLYII	NPANKGYLVE	TDPFRFANYRQ	WIGSDYMLGS
151	LKLDPNLHK	RLGDSYEQRL	LINEQIAELT	GHRRLDGYQN	DEEQFKALMD
201	NGATAARSMN	LSVGIALSAE	QAAQLTSDIV	WLVOKEVKLP	DGGTQTVLVLP
251	QYVYVRVKNNG	IDGKGALLSG	SNTQINVSGS	LKNSGTIAGR	NALIINTDTL
301	DNIGGRIHAQ	KSVAVTATQDI	NNIGGILSAE	QTLNAGNN	INNSQSTAKSS
351	QNAQGSSTYL	DRMAGIYITG	KEKGVLAQA	QKIDNIIAGQ	ISNSQDQGT
401	RLQAGRDIND	DTVQGTGYQE	IEFADNHTIR	IGSTNVEGSS	IQTKGVDVLL
451	SGNNLAKRA	EYCSAKETLA	VYAKNDITIS	SGIHAGQYDD	ASKHTRSGSG
501	GNLVIYTDKA	QSHHETQSS	TEEGKQVVLQ	AGNDANILGS	NVISDNHTRI
551	QAGNHVIRIGT	QTQTSQSEY	HTQKSGIMS	AGIGFTIGSK	NTQENQSSQS
601	NEHTGSTVGS	LKGDITTVAS	KHYEQTGSNV	SSPFGNNLIS	TQSMOIGAAQ
651	NQINSKTTQT	YEQKGLTVAF	SSPVTDLAQQ	AIWAHAKKAK	QFDKATTTAL
701	MPWRLFMQVG	RLFKQAKAPK	K*		

Further work revealed the following partial gonococcal DNA sequence <SEQ ID 515>:

1	TTGCTTGTC	AAACAGAAAA	AGACGGTTTG	CATAACGAGC	AAACCTTTGG
51	CGAGAGAGAA	GTCTTCAGCG	AAATGTTAA	GTTCACACAC	TACTGGCGTG
101	CGCGTGTAA	AGGACATGAT	GAACAGGGC	ATCGTGACAT	AAATTTACT
151	TTGCGCGAGG	AAATCACACG	CGACATTTC	CTGGGTTTCT	TTGCCTATGA
201	ATCGCATAGC	AAAGCATTA	CGCGTCATGC	GCCGACGCCA	GCGACTGAGT
251	TGCCACAAG	TAACCGGGAT	AATATCCGTA	CTGCGAAGAG	CAACGGTATT

301	TGCGTACCGT	ATAGCGCCAA	TTCTTTTACC	CCATTACCGG	GCAGCAGCTT
351	ATACGTTATC	ATACGTTATC	ATACGTTATC	ATACGTTATC	ATACGTTATC
401	GCTTTGGCAA	CTACCGTCAA	TGTTTGGGTA	CTGACTATAT	GCTGGGCGAC
451	CTCAAACTAG	ACCCAAACAA	TTTACATAAA	CGTTTGGGTG	ATGTTTATTA
501	CGAGCAACGT	TTAATCAATG	AACAAATCGC	AGAGCTGACA	GGGCACTCGT
551	GTTTAGAAGG	TTATCAAAAC	GACGAAAGAC	AATTTAAAGC	CTTAATGGAT
601	AATGGCGCGA	CTGCGGCACG	TTGATGAAT	CTCAGCGTTG	GCATTGTCATT
651	AAGTGCGCGA	CAGCAGCGCG	AACGTACACG	CGATATTGTT	TGGTTGGTAC
701	AAAAAAGAGT	TAAACTTCTT	GATGGCGGCA	CACAAACCGT	ATTGATGCCA
751	CAGGTTTATG	TACGCGTTAA	AAATGGCGGC	ATAGACCGTA	AAGGTGCATT
801	GTGTGACGGC	AGCAATACAC	AAATCAATGT	TTACGAGCAG	CTGAAAAACT
851	CAGGCACGAT	TGCAGGCGCG	AATGCGCTTA	TTATCAATAC	CGATACGCTA
901	GACAATATCG	GTGGGCGTAT	TCATGCGCAA	AAATCAGCGG	TTACGGCCAC
951	ACAAGACATC	AATAATATTG	GCGGCATTCT	TTCTGCGCAA	CAGACATTAT
1001	TGCTCAATGC	GGGTAAACAC	ATCAACAACC	AAAGCAGCGG	CAGAGCAGST
1051	CAAAATGCAC	AAGGTAGCAG	CACCTACCTA	GACCGAATGG	CAGGTATTTA
1101	TATCAGCGGC	AAAGAAAAGG	GTGTTTATAG	AGCGCAGGCA	GGCAAGAGCA
1151	TCAACATCAT	TGCGGTCGAA	ATCAGCAATC	AATCAGATAT	AGGCAAAACC
1201	CGGCTCGAGG	CAGGACGCGA	CATTTAACTG	GATACGGTAC	AACCGGCAAA
1251	ATATCAAGAA	ATCCATTTTG	ATGCGGATAA	CCATACCATC	CGAGGTTCAA
1301	CGAACGAAGT	CGGACGACAG	ATTCAAACAA	AAGGCGATGT	TACCCtatTG
1351	TCAGGGAATA	ATCTCAATGC	CAAGGCTGCC	GAAGTCGGCA	GCGCCAAAGG
1401	CACACTTGCC	GTGTATGCTA	AAAATGRCAT	TACTATCAGC	TCAGGCATCC
1451	ATGCGCGGCA	AGTTGATGAT	CGGTCACAA	ATACAGGACG	AAGCGCGCGC
1501	GGTAATAAAT	TAGTCATTAC	CGATAAAGCC	CAAAGTCATC	ACGAAACTGC
1551	TCAAAGCAGC	ACCTTTGAAG	GCAAGCAAGT	TGTATTGCAG	GCAGGAAAGT
1601	ATGCCAACAT	CTTGGCAGT	AATGTTATT	CCGATTAATG	CACCCGGATT
1651	CAAGCAGGCA	ATCATGTTTG	CATTGGTACA	ACCCAAACTC	AAAGCCAAAG
1701	CGAAACCTAT	CTCAAAACCC	AAAAATCAGG	ATTGATGAGT	GCAGGTATCG
1751	GCTTCACTAT	TGGCAGCAAG	ACAAACACAC	AAGAAACCAA	ATCCCAAAGC
1801	AACGAACATA	CAGGCAGTAG	CGTAGGCAGC	CTGAAAGGCG	ATACCACTAT
1851	TGTTGCAAGC	AAACACTACG	AACAAACCGG	CAGCAAGCTT	TCACGCCCTG
1901	AGGCCAACAA	CTTTATCAGC	ACGCCAAACC	TGGATATTGG	CGCAGCACAA
1951	AACCAATATC	CTTTATCAGC	CACCCAAACC	TAGCAACAAA	AAGGCTTAAC
2001	GCTGGCATTG	AGTTGCGCGC	TTACCGATTG	GGCACAACAA	GGGATTCGCG
2051	TAGCACACAA	AGCAACAAAC	AACTCGGACA	AGCAAAAAC	AACCGCGTTA
2101	ATGCCATGCG	GGCTGCCAAT	CGAGTTTGGC	AGGCTATACA	AACAGGCCAA
2151	GGCGACAA	ACTTAG			

This corresponds to the amino acid sequence <SEQ ID 516; ORF115ng-1>:

40	1	LLVQTEKDGL	HNEQTFGEKK	VFSENGKLHN	YWRARRRGHD	ETGHRQNYT
	51	LPPEITRDIS	LGSFAYESH	KALSRHAPSQ	GTELPQSNRD	NIRTKASNGI
	101	SLPYTPNSFT	PLPGSSLYII	NPANKGYLVE	TDPRFANYRQ	WLGS DYMLGS
	151	LKLDPNNLHK	RLGDGYYEQR	LINEQIAELT	GHRRLDGYQN	DEEQFKALMD
	201	NGATAARSMN	LSVGLIALSE	QAQLTSDIV	WLVKGEVKPL	DGGTQVLMPL
45	251	QVYVVRKNGG	IDKGALLSG	SNVTQNVSGS	LKNSGTIAGR	NALIINTDTL
	301	DNIGRIHAQ	KSAVTAQDI	NNIGGILSAE	QTLNLNAGNN	INNQSSTAKSS
	351	QNAQSSSTYL	DRMAGIYITG	KEKGVLAQA	GKIDINIAGQ	ISNQSDQGT
	401	RLQAGRDINL	DTVTQKYQE	IHPDADNHTI	RGSTNEVGSS	IQTGKDVTL
50	451	SGNNLNKAA	EVGSAKGTLA	VYAKNDITIS	SGIHAGQVDD	ASKRTGRSGG
	501	GNKLVTDKA	QSHHETAGSS	TFEGKQVVLQ	AGNDANILGS	NVISDNGTRI
	551	QNHVHRLCT	TYTQSQSETH	KQTKQGLMS	AGIGFTIGSK	TNTQENQSGS
	601	NEHTGSTVGS	LKGDITLIVAS	KHVEGTSSNV	SSPFTGILLIS	TGSMIDIGAAQ
	651	NQLMSKTQQT	YEQKGLTVAE	SSPVDTLAQQ	ATAVAHKAAN	KSDKAKTATL
	701	MPWRLMPQVG	RP:KQAKAHK	T*		

55 This gonococcal protein (ORF115ng-1) shows 91.9% identity with ORF115 over 334aa:

		20	30	40	50	60	70
	orf115ng-1.p	NEQTFGEKKVFSENGKLHN	YWRARRRGHDETGHRQNYTL	PEEITRDISLGSFAYESHK			
	orf115			STGHSEQNYTLPREITRNISLGSFAYESHK			
60				10	20	30	
		80	90	100	110	120	130
	orf115ng-1.p	ALSRHAPSQGTLEPQSNRDNIR	TAKSNGISLPTYTPNSFTPLPGSSLYII	NPANKGYLVET			
65	orf115	ALSHHAPSQGTLEPQSN	-----	GISLPTYTPNSFTPLPGSSLYII	NPANKGYLVET		
		40	50	60	70	80	

		140	150	160	170	180	190
	orf115ng-1.p	DPRFANYRQWLGSDYMLGSLKLDPNNLHKLRLGSGYEQRLINEQIAELTGHRRLDGYQND					
5	orf115	DPRFANYRQWLGSDYMLGSLKLDPNNLHKLRLGSGYEQRLINEQIAELTGHRRLDGYQND					
		90	100	110	120	130	140
	orf115ng-1.p	EEQFKALMDNGATAARSMNLSVGLIALSAFQAQALTSDIVLWVQKEVKLPDGGTQTVLMPQ					
10	orf115	EEQFKALMDNGATAARSMNLSVGLIALSAEQVAQLTSDIVLWVQKEVKLPDGGTQTVLMPQ					
		150	160	170	180	190	200
	orf115ng-1.p	VYVRVKNKGIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDITLNDIGGRIHAQK					
15	orf115	VYVRVKNKGIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDITLNDIGGRIHAQK					
		210	220	230	240	250	260
	orf115ng-1.p	SAVTATQDINNIGGILSAEQTLNAGNNINNSQSTAKSSQNAQGSSTYLDORMAGIYITGK					
20	orf115	SAVTATQDINNIGGILSAEQTLNAGNNINNSQSTASSONTQGSSTYLDORMAGIYITGK					
		270	280	290	300	310	320
25	orf115ng-1.p	EKGVLAAQAGKDINIIAGQISNSDQGGTQLQAGRDINLDTVQTKGYQEIHFADNHTIR					
	orf115	EKGVL					

In addition, it shows homology with a secreted *N.meningitidis* protein in the database:

30	gi 2623258 (AF030941) putative secreted protein [Neisseria meningitidis] Length = 2273
	Score = 604 bits (1541), Expect = e-172
	Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)
35	Query: 1 LIVGTTEKGDHNEQTFGEKKVFSENGKLHNYWRARRKGHDETHREQNTYLPEEITRDIS 60
	I+V T + L N++T G K K + ++ G L R Y R + K G D T C + Y E + + I
	Sbjct: 739 LIVGTPEALNDNETLGTCTI-TDKGDLHRYHRHHKKGRDSTGYRSRPFYPAFEVS-SIR 796
40	Query: 61 LGSFAYESHKALSRHAPSGQTELPQSNRDNIRTAQNSIGSLPYTPNSFTPLPGSSLYII 120
	+G AY+ + AP Q +++P + + NGI +T LP SSL+ I
	Sbjct: 797 MGISAYKGY-----APQASDIPGTV---VPVVAENGHPTFT-----LNSSSLFAI 840
45	Query: 121 NFANKGYLVETDPRFANYRQWLGSDYMLGSLKLDPNNLHKLRLGSGYEQRLINEQIAELT 180
	P NKGYL+ETDF P +YR+WLGS YML +L+ DFN++HKRLGSGYEQ+L+NEQIA+LT
	Sbjct: 841 APNNKGYLIETDPAFTDYRKWLGSYMLAALQDDPNHHLKRLGSGYEQKLVNEQIAKIT 900
	Query: 181 GHRRLDGYONDEEQFKALMDNGATAARSMNLSVGLIALSAEQVAQLTSDIVLWVQKEVKLP 240
	G+RRLDGY NDEEQFKALMDNG T A+ +L+ GIALSAEQ A+LTSDIVWL + V LP
	Sbjct: 901 GYRRLDGYTNDEEQFKALMDNGITIAKEQLTPEGIALSAEQVARITSDIVWLENETVTLP 960
50	Query: 241 DGGTQTVLMPQVYVRVKNKGIDGKGALLSGSNTQINVSGSLKN-SGTIAGRNALIINTDT 299
	DG TQTVL P+VYVR + ++GALLSGS I SG+++N G IAGR ALI+N
	Sbjct: 961 DGGTQTVLKPQVYVRARPKDMNGQALLSGSVVDIG-SGAIENRGLIAGREALILNAQN 1019
55	Query: 300 LDNIGGRIHAQKSAVTATQDINNIGGILSAEQTLNAGNNINNSQSTAKSSQNAQGSSTY 359
	+ N+ G + + A DI N G I AE LLL A NNI ++S +S+QN QGS
	Sbjct: 1020 IKNIQSGDLQGNIFAAAGSDITNTGSI- GAENALLKASNIERSRSTRSNQNEQGSVRN 1078
	Query: 360 LDRMAGIYITGKEKGVLAQAQAGKDINIIAGQISNSDQGGTQLQAGRDINLDTVQTKGYQ 419
	+ R+AGIY+TG++ G + AG +I + A +++NQS+ GGT L AG DI DT + Q
60	Sbjct: 1079 IGRVAGIYLTGRQNGSVLLDAGNNIVITASELTNQSEDGQTVLNAGGDIISDPTGISRQ 1138
	Query: 420 EHFADNHTIRGSTNEVGSSIQTGKDVTLTSGNNLNAKAAEVGSAKGTAVAYAKNDITI 479
	FD+NDY IR NEVGSI+T+G+++L + ++ +AAEVGS +G L + A DI +
65	Sbjct: 1139 NTIFDNYVIRKQNEVGSTIRTRGNLSNAGKDIRIAAEVSGEQGRKLAAAGRIK 1198
	Query: 480 SSGIHAGGVDDASKHTGRSGGKGLVITDKAQSHHETAQSSSTFEKGQVVLQAGNDANILG 539
	+G +D+K+TGRSGG K +T + + + A S T +GK++ +G D + G
	Sbjct: 1199 EAGKAHTEDEALKYTGSRGGGKIKQMRHLKNQNGQAVSGTLDGKEIILVSGRDIITVG 1258

Query: 540 SNVISDNGTRIOAGNHVRIGTTQTOSQSEYHQTOKSLGM-SAGIGFTIGSKTNTQENQS 598
 SN+I+DN T + A N++ + +T+S+S ++ +KSGLM S GIGFT GSK +TQ N+S
 Sbjct: 1259 SNIADNHTILSAKNNIVLKAAETRSRAEMNKKEKSGLMSSGIGFTAGSKDTQTNRS 1318

5 Query: 599 QSNHETGSTVSGSLKGDTTIVASKHYEQTGSNNVSSPEGNLLISTQSMIDGAAQNIQNSKT 658
 ++ HT S VGS L G+T I A KHY QTGS +SSP+G+ IS+ + I AAQN+ ++
 Sbjct: 1319 ETVSHTESVSGSLNGNTLISAGKHYYTQTGSTISSPQGDVGIISSGKISIDAAQNRYSQESK 1378

10 Query: 659 QTYEQKGLTVAFSSPVT 676
 Q YEOKG+TVA S PV +
 Sbjct: 1379 QVYEOKGVTVAISVPVN 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 62

15 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 517>:

1 ..TCAGGGAATA ACCTCAATGC CAAAGCTGCC GAAAGTCAGA GCGCAAACGG
 51 TACACTCGCT GTGCTCGCCA ATATGACAT CAACATCAGC GCAGGCATCA
 101 ACACGACCCA TGTGTATGAT GCGTCCAAAC ACACAGGCAG GAGCGGTGGT
 151 GGCAATAAAT TAGTCATATC CGATAAAGCC CAAAGTCATC ACGAAACCCG
 201 CCAAGCAGC ACCTTTGAG CCAAGCAAGT TGTATTGCGC GCAGGAAACG
 251 ATGCCAACAT CCTTGGCAGC AATGTTATTT CCGATAATGC CACCCAGATT
 301 CAAGCAGGCA ATCATGTTCC CATTGGTACA ACCCAAACCT AAAGCCAAAG
 351 CGAAACCTAT CATCAAACCC AGAAATCAGG ATTGATGAGT GCAGGTATCG
 401 GCTTCACTAT TGGCAGCAGC ACACACACAC AAGAAAACCA ATCCCAAAGC
 451 AACGAACATA CAGGCAGTAC CGTAGGCAGC TTGAAGCGC ATACCAACAT
 501 TGTTCAGGCG AAACACTACG AACAAATCGG CAGTACCGTT TCCAGCCCGG
 551 AAGGCACACA TACCATCTAT GCCCAAGACA TAGACATTTA AGCGGCACAG
 601 AACAAATTAA ACATAATATC CACCCAAACC TATGACACAA AAGG. CTAAC
 651 GTTGGCATTC AGTTCGCCCG TTACCGATTI GGCACACAAA ...

30 This corresponds to the amino acid sequence <SEQ ID 518; ORF117>:

1 ..SGNNLNAAKAA EVSSANGTLA VSAANDINIS AGINTHVDV ASKHTGRSGG
 51 GNKLIVITDKA QSHHETAQSS TFEKQVQLV AGNDANILGS NVISDNGTQI
 101 QAGNHVRIGT TOTOSQSEY HQTOKSLGMS AGIGFTIGSK TQENQSQS
 151 NEHTGSTVGS LKGDTTIVAG KHYEQIGSTV SSPFNNTIY AQSIDIQAAH
 201 NKLSNNTQT YEOKXLTVAF SSPVTLAQ ...

Computer analysis of this amino acid sequence gave the following results:

Homology with the pspA putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF117 and pspA protein show 45% aa identity in 224aa overlap:

Orf117: 4 NLNAKAEVSSANGTLAVSANNDINISAGINTHVDVASKHTGRSGGKNLVIITDKAQSH 63
 ++ +AAEV S G L ++A DI + AG T +DA E+TGRSGGG K +T ++
 pspA: 1173 DIRIRAAEVGSEQRLKLAAGRDIKVEAGKAHTEDEALRYTGRSGGKQKMTRHLLKXQ 1232

Orf117: 64 HETAQSSTFEGKQVQLVAGNDANILGSNVISDNGTQIQAGNHVRIGTTQTOSQSEYHQT 123
 + A S T +GK+++L +G D + GSN+I+DN T + A N++ + +T+S+S ++
 45 pspA: 1233 NGQAVSGTLDGKEIILVSGRDIPTVTSNIIADNHTILSAKNNIVLKAAETRSRAEMNKK 1292

Orf117: 124 KKSGLM-SAGIGFTIGSKTNTQENQSQSNHETGSTVSGSLKGDTTIVAGKHYYEIGSTVSS 182
 +KSGLM S GIGFT GSK +TQ N+S++ HT S VGS L G+T I AKHY Q GST+SS
 pspA: 1293 EKSGLMSGGIGFTAGSKDTQTNRETSVSHTESVGSGLNGNTLISAGKHYYTQTGSTISS 1352

50 Orf117: 183 PEGNNTIIYAQSIDIIQAHHKNSNTTQTYEQKXLTVAFSSPVT 226
 P+G+ I + I I AA N++ + + Q YEOK +TVA S PV +
 pspA: 1353 PQGDVGIISSGKISIDAAQNRYSQESKQVYEOKGVTVAISVPVN 1396

Homology with a predicted ORF from *N. gonorrhoeae*

ORF117 shows 90% identity over a 230aa overlap with a predicted ORF (ORF117ng) from

N. gonorrhoeae:

5	orf117.pep	SGNNLNAAAEVSSANGTLAVSANNNDINIS	30
	orf117ng	IHFADNHTIRGSTNEVGSSIQTKGDVTLTSGNNLNAAAEVSSAGTKLAVYAKNDITIS	480
	orf117.pep	AGINTTHVDDASKHTGRSGGGNKLIVITDKAQSHHETAQSSTFEKGQVVLQAGNDANILGS	90
10	orf117ng	SGIHAGQVDDASKHTGRSGGGNKLIVITDKAQSHHETAQSSTFEKGQVVLQAGNDANILGS	540
	orf117.pep	NVISDNGTQIQAGNHVRIGTTQTQSQSETHYQTKQSGLSAGIGFTTIGSKTNTQENQSQS	150
15	orf117ng	NVISDNGTQIQAGNHVRIGTTQTQSQSETHYQTKQSGLSAGIGFTTIGSKTNTQENQSQS	600
	orf117.pep	NEHTGSTVGSCLKGDDTIVAGKHVEQIGSTVSSPEGNNTIYAQSIDIAAHNNKLSNNTQT	210
	orf117ng	NEHTGSTVGSCLKGDDTIVAGKHVEQIGSTVSSPEGNNTIYAQSIDIAAHNNKLSNNTQT	660
20	orf117.pep	YEQKLTVAFFSSPVTDLAQQ	230
	orf117ng	YEQKLTVAFFSSPVTDLAQQAIAVAHKAARKQFDKAKTTALMPWRLPMQVGRLFQKAKAPK	720

An ORF117ng nucleotide sequence <SEQ ID 519> was predicted to encode a protein having amino acid sequence <SEQ ID 520>:

25	1	..LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETCHREQNTY
	51	LPEEITRDIS LGSFAYESH KALSRAHFSQ GTELPQSNRD NIRTAKSNGI
	101	SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPRFANYRQ WLGS DYMLGS
	151	LKLDPNNLHK RLGDGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
30	201	NGATAARSMN LSGGIALSAE QAAQLTSDIV WLQVEKVLK DGGTQTVLMP
	251	QYVVRVKNKG IDHGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
	301	DNIGGRIHAQ KSAVATQDI NNIGGTLSE QTLNLAGN NNQSTAKSS
	351	QNAQGSSTYL DMAGIYITG KEGVLAQAQ GKDINI IAGQ ISNQSDTAKS
	401	RLQAGRDNL DVTGQKQE IHFDNHTI RGSNEVGSS LQTKGDVTL
35	451	SGNNLNAAAEVSSAGTKLAVYAKNDITIS
	501	GKLVITDKA QSHHETAQSS TFEKGQVVLQ AGNDANILGS NVISDNGTRI
	551	QAGNHVRIGT TQTQSQSETHYQTKQSGLS AGIGTTIGSK TINTQENQSQS
	601	NEHTGSTVGS LKGDDTIVAS KHVEQIGSNV SSPEGNNTIS TQSMIDIAAQ
	651	NQLSKTTQT YEQKLTVAFFSSPVTDLAQQ AIAVAHKAARK QFDKAKTTAL
	701	MPWRLPMQVG RLFQKAKAPK K*

40 Further work revealed the following gonococcal partial DNA sequence <SEQ ID 521>:

	1	TTGCTTGTC AAACAGMAAA AGACGGTTTG CATACGAGC AACCTTTGG
	51	CGAGAAGAAA GTCTTCAGCG AAAATGGTAA GTTGACACAC TACTGGCGTG
	101	CGCGTCGTAA AGGACATGAT GAACACGGCG ATCGTGACAA AATATATAC
45	151	TTGCCGCGAGG AATACACACG CGACATTTCAT CTGGGTTCAT TTGCCTATGA
	201	ATCGCATAGC AAAGCATTAA CGCGTCATGC GCCACGCCAA GGCCTAGCT
	251	TGCCACAAAG TAACCGGGAT AATATCGGTA CTGCGAAAG CAACGCTATT
	301	TGCTACCTC ATACGCCCAA TTCTTTTACC CCAATACCG CGACGAGCTT
	351	ATACATTATC AATCCTGCCA ATAAAGGCTA TCTTGTGTAA ACCGATCCAC
50	401	GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACATAT GCTGGGCGAG
	451	CTCAACCTAG ACCCAACAA TTTACATAAA CGTTTGGGTT ATGGTTATTA
	501	CGGCGACCT TTATCATCG ACACATCGC AGGCTTGACA GGCGATCGCT
	551	GTTTAGACGG TTTACAAAC GACGAAGAC AATTATGAC CTATATGAT
	601	AATGGCGCGA CTGCGCGACG TTGATGAAT CTACGCTTG CATTCGATT
	651	AAGTGCGGAG CACACGCGC AACTGACGAG CGATATTGTT TGGTTGGTAC
55	701	AAAAAGAGT TAACTTCCT GATGCGCGCA CACAAACCGT ATTGATGCCA
	751	CAGGTTTATG TACGCGTTAA AATGCGCGC ATAGACGGTA AAGTGCAATT
	801	GTTGTCAGCG AGCAATACAC AATCATATGT TTAGCGCAGC CTGAAAAACT
	851	CAGGACGAGT TGCAGGCGCG AATGCGCTTA TTATCAATAC CGATACGCTA
	901	GACAATATCG GTGGCGGTAT TCATGCGCAA AATACGCGG TTACGCGCAC
60	951	ACAAGACATC AATAATATG CGGCGATTCT TTCTCGGAA CAGACATTAT
	1001	TGCTCAATGC GGTAAACAC ATCAACACC AAGCAGCGC CAAGAGCATG
	1051	CAAAATGCAC AAGGTAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA

1101	TATCACAGGC	AAAGAAAAG	GTGTTT'AGC	AGCGCAGGCA	GGCAAAGACA
1151	TCAACATCAT	TGCCGCTCAA	ATCAGCAATC	AATCAGATCA	AGGCGCAACG
1201	CGGCTCGCAG	CAGGACGCGA	CATTAACTCT	GATACGGTAC	AAACCGGCAA
1251	ATATCAAGAA	ATCCATT'TTG	ATGCCGATAA	CCATACCATC	CGAGGT'TCAA
1301	CGAACGAAGT	CGGCAGCAGC	ATTCRAACAA	AAGGCGATGT	TACCC'atATG
1351	TCAGGGAATA	ATCTCAATGC	CAAAAGT'GCC	GAAGTCGGCA	GGCGAAAAGG
1401	CACACTTGCC	GTGTATGCTA	AAAATGACAT	TACTATCAGC	TCAGGCATCC
1451	ATGCCGCGCA	AGTTGATGAT	CGGTCCAAAC	ATACAGGCAG	AAGCGCGCCG
1501	GGTAATAAAT	TAGTCAATAC	CGATAAAGCC	CAAAAGT'CATC	ACGAAACTCG
1551	TCAAAGCAGC	ACCTTTGAAG	GCAAGCAAGT	TGTATTGACG	GCAGGAAGCG
1601	ATGCCAACAT	CCTTGGCAGT	AATGTTATTT	CCGATAATGG	CACCGGGATT
1651	CAAGCAGGCA	ATCATGTTTC	CATTGGTACA	ACCCAAACTC	AAAGCCAAAG
1701	CGAAACCTAT	CATCAAAACC	AAAAATCAGG	ATTGATGAGT	GCAGGTATCG
1751	GCTTCACTAT	TGGCAGCAGG	ACAAACACAC	AAGAAAACCA	ATCCCAAAGC
1801	AACGAACATA	CAGGCAGTAC	CGTAGGCAGC	CTGAAAGGCG	ATACCAACAT
1851	TGTTGCACAG	AAACACTACG	AACAACCGG	CAGCAAGCTT	TCCAGCCCTG
1901	AGGCGACAAA	CCTTATCAGC	ACGCAAAAGT	TGGATATTGG	CGACGCACAA
1951	AACCAATTAA	ACACAAAAAC	CACCCAAAAC	TACGAACAAA	AAGCGTTAAC
2001	GGTGGCATTC	ACTTCGCGCG	TTACCGATTT	GGCACAACAA	GGATTGCGG
2051	TAGCACACAA	AGCAGCAAAC	AAGTCSGACA	AAGCAAAACG	GACCGCGTTA
2101	ATGCCATGGC	GGCTGCCAAT	GCAGTTGGC	AGGCGTATCA	AACAGGCAAA
2151	GGCGCACAAA	ACTTAG			

This corresponds to the amino acid sequence <SEQ ID 522; ORF117ng-1>:

25	1	LLVQTEKDG	L	HNEQTFGEKK	V	FSENGLKLN	Y	WRARRKGHD	E	THGREQNYT
	51	LPEEITRDIS	L	GSFAYESH	S	KALSRHAPSQ	G	TGELPQSNRD	N	IRTKASNGI
	101	SLPYTPNSFT	P	LPSSLYII	N	PANKGYLVE	T	DPFRFANYRQ	W	LSGSDYMLGS
	151	LKLDPNLHKL	R	LDGDDYEQ	R	LINEQIAELT	G	HRRLDGYQN	D	EEQFKALMD
	201	NGATAARSMN	L	SVGIALSAE	A	QAALQTSIDV	W	LVQKEVKLF	D	GGTQTVLMP
	251	QVYVRVKGNG	I	DGKALLSG	S	NTQINVS	G	LKNSGT IAGR	N	ALIINTDTL
30	301	DNIGGRIHAQ	K	SAVTATQDI	N	NIGGILSAE	Q	TLLLNAGNN	I	NNSTAKSS
	351	QNAQGSSTYL	D	RMAGIYITG	K	ERGVLAQAQ	G	KDINI IAGQ	I	SNOSDQGT
	401	RLQAGRDINL	D	TVGTGKYQE	I	HFADNHTT	R	GSSTVSGSS	I	QTKGDTVLL
	451	SGNLLNARAA	E	VCSAGKTLA	V	AKNDITIS	G	GHAGQVDD	A	SKHTIRASGG
	501	SNKLVITDKA	O	SHHEITAGS	T	EEGQVQLQ	A	NDANLILGS	N	VLSDQKTLI
35	551	QAGNHVIRGT	T	CTOSQSETY	H	OTKCSGLMS	A	IGITPTTSGK	T	NTQENQSGS
	601	NEHTGSTVGS	L	KGDITIVAS	K	HYEQTGSNV	S	SPSENNLIS	T	QSMIDGAQ
	651	NQLNSKTTQT	Y	EQKGLTVAF	S	SFPTDLAQQ	A	IAVAHKAAN	K	SKDKARTAL
	701	MPWRLEPMQVG	R	PIKQAKAHK	T					

ORF117ng-1 shows the same 90% identity over a 230aa overlap with ORF117. In addition, it shows homology with a secreted *N.meningitidis* protein in the database:

	g12623258 (AF030941) putative secreted protein [Neisseria meningitidis] Length = 2273
	Score = 604 bits (1541), Expect = e-172
	Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)
45	Query: 1 LLVQTEKDLHNEQTFGEKKVFSENGLKLNHYWRARRKGHDETHGREQNYTLP EETITRDIS 60
	L-V T + L N++T G K + ++ G L H Y R +KG D TG+ Y E++ I
	Sbjct: 739 LVGTPEESALNDDET LTKTII-TDKGLDHRHYRHHKGRDSTGYSRSPYEPAP EVS-SIR 796
50	Query: 61 LGSFAYESHSKALSRHAPSQGTGELPQSNRDNIRTKASNGISLPYTPNSFTPLPGSSLYII 120
	+G AY+ + AP Q ++++ + + NGI +T LP SSL+ I
	Sbjct: 797 MGISAYKGY-----APQASDIPGTV--VPVVAENGHIPTFT-----LPNSSLFAT 840
55	Query: 121 NPANKGYLVETDPRFANYRQWLGSDDYMLGSLKLDPNLHKLRLDGYEQRLINEQIAELT 180
	P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKLRLDGYEQ+L+NEQIR+LT
	Sbjct: 841 APNKGYLITETDPAFTDYRKWLGSGLMIAALQDDPNHIIHKLRLDGYEQYKQLVNEQIAELT 900
	Query: 181 GHRRLDGYONDEEQFKALMDNGATAARSMNLSVGIALSAEQAAALQTSIDVWLVOKEVKLP 240
	G+RLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQ A+LTSIDVWL + V LP
60	Sbjct: 901 GYRLDGYTNDDEEQFKALMDNGITIAKELQTPGIALSAEQVARLTSIDVWLNEETVETLP 960
	Query: 241 DGGTQTVLMPQVYVRVKGNGIDGKALLSGSNTQINVSGLKN-SGTIAGRNALIINTDT 299
	DG TQTVL P+VYVR + ++GALLSGS I SG++N G IAGR ALI+N
	Sbjct: 961 DGGTQTVLKPQVYVRARPKMNGGALLSGSVVDIG-SGAENRGGLIAGRREALIINQNN 1019
65	Query: 300 LDNIGGRIHAQKSAVTATQDINNIGGILSAEQTLNLLNAGNNINNSTAKSSQNAQGSSTY 359

-307-

+ N+ G + + A D I N G I A E L L L A N N I + S + S + Q N Q G S
 Sbjct: 1020 IKNLQGDILQGNIFAAAGSDITNTGSI-GAENALLKASNNIESRSSETRSNQNEQSSVRN 1078
 Query: 360 LDRMAGIYITGKEKGVLAQAQKDNIIAGQISNQSDQOQTRLQAGRDINLDTVTQTKYQ 419
 + R+AGIY+TG++ G + AG + I + A +++NQS+ GQT L AG DI DT + Q
 5 Sbjct: 1079 IGRVAGIYLTGRQNSVLLDAGNNIVLTASELTNQSEDOQTVLNAGGDIRSDTGTISRN 1138
 Query: 420 EIHFDADNHTIRGSTNEVGSSIQTKGDVTLTLLSGNNLNKAAAEVGSAGKTLAVYAKNDITI 479
 FD+DN+ IR NEVGS+I+T+G+++L + ++ +AAEVGS +G L + A DI +
 10 Sbjct: 1139 NTFIDSDNYVIRKEQNEVGSTIRTRGNLSLNKAGDRIIRAEEVSGEGGRKLKLAAGRDIKV 1198
 Query: 480 SSGIHAGQVDDASKHTGRSGGSKLVIITDKAQSHHETAQSSTFEKGQVVLQAGNDANILG 539
 +G + +DA K+TRSGGS K +T + + A S T +GK+++L +G D + G
 15 Sbjct: 1199 EAGKAHTETEDA LKYTRSGSGGKIKQKMRHLKNQNGQAVSGTLDGKEIILVSGRDTITVG 1258
 Query: 540 SNVISDNMTRIQAGNHVRIGTITQTQSQSEYHQTKSGLM-SAGIGFTIGSKNTNTQENQS 598
 SN+I+DN T + A N++ + +T+S+S ++ +KSGLM S GIGFT GSK +TQ N+S
 Sbjct: 1259 SNIIADNHTIILSAKNNIVLKAETRSRAEMNKKEKSGLMGSGGIGFTAGSKKDTQNR 1318
 Query: 599 QSNNEHTGSTVGSLEKGDITTVASKHYEQTGSNVSSPEGNILISTQMSMDIGAAQNLQSKT 658
 ++ HT S VGS L G+T I A KHY QTGS +SSP+G+ IS+ + I AAGN+ + + +
 20 Sbjct: 1319 ETVSHTESVVGSLNGNTLISAGKHITQTGSTISSPQGDVGISSGKISIDAAQNRYSQESK 1378
 Query: 659 QTYEQKGLTVAFSSPVT 676
 Q YEQKG+TVA S PV +
 25 Sbjct: 1379 QVYEQKGVTVAISSPVVN 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 63

30 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 523>:

1 ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCTCGCGG TTGTGCGCTA
 51 CACATGTTAT CAGGAAAACC AATACCGCAA AATAGTGGCG GACCAATTCCG
 101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCWAACCCAG CCAATGTCGG
 151 GACGGCAAAAC CGTCCGGCGG GTCACTCATG ATCGCGAAAC CCCAACCGCG
 35 201 GGTCAAAAAA ACGGCAAAAC CCCAAGAACC CGYCATGCC AACCTGCAAG
 251 AACAGGATCC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG
 301 TTCAAAACCG AATCGAAAC CGCCTTGAA GAAGCGGCA TTATCGGCAA
 351 CTCGCCCAAC ACCGTTTCCG AACCCCAAAC CGGACATTCC GCAACGAAAC
 401 CTGCGGACGC GTGCGCAAAA CTTGCAACCG TTCCGCAAA ACCTGCAAAA
 40 451 CGCGTGATTA CGCTCAAGA ACTGTCAAAA GTCGAATTAT CCTGGTTTGA
 501 CGTGGCATC GACTTCATCT CCTAT...

This corresponds to the amino acid sequence <SEQ ID 524; ORF119>:

1 MIYIVLFLAV VLAVVAYNMY QENQYRKVR DQFHSKDKA LLNSXTSRVR
 51 DGKPSGGSM MPKQPAPVK TAKPQDPXMR NLQEQDAVYI AKQKQAKASP
 45 101 FKTEIETALE ESLIIGNSAH TVSEPQTGHS ATKPADASAK PAPVPQTPAK
 151 PLITLKEISK VELSWDFVRI DFISY...

Further work revealed the complete nucleotide sequence <SEQ ID 525>:

1 ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCTCGCGG TTGTGCGCTA
 51 CACATGTTAT CAGGAAAACC AATACCGCAA AATAGTGGCG GACCAATTCCG
 50 191 GACACTCCGA CAAAGATGCC CTGCTCAACA GCWAACCCAG CCAATGTCGG
 151 GACGGCAAAAC CGTCCGGCGG GTCACTCATG ATCGCGAAAC CCCAACCGCG
 201 GGTCAAAAAA ACGGCAAAAC CCCAAGAACC CGYCATGCC AACCTGCAAG
 251 AACAGGATCC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG
 301 TTCAAAACCG AATCGAAAC CGCCTTGAA GAAGCGGCA TTATCGGCAA
 55 351 CTCGCCCAAC ACCGTTTCCG AACCCCAAAC CGGACATTCC GCAACGAAAC
 401 CTGCGGACGC GTGCGCAAAA CTTGCAACCG TTCCGCAAA ACCTGCAAAA
 451 CGCGTGATTA CGCTCAAGA ACTGTCAAAA GTCGAATTAT CCTGGTTTGA
 501 CGTGGCATC GACTTCATCT CCTATATGCG GTGACCGAA GCCAANGAAC
 551 TGCAACGACT GCGCGGCGCT TCACAACCGCT GCGGCTACCA GATTGTCCGC
 60 601 TGCAACCATG ACGACCATTT CCAGATTGCG GAACCATCC CGGCGATCCG

5 651 CTATCAGGCA TTTATCGTGG GTATTTCAGGC AGTCAGCGCG AAGCGACTTG
 701 CCTCCAGGA AGAAGCTCTG CATTTCAGC GCAGGTGCGA GCATTCGCA
 751 CAACACATG GCGCTCAGAC GCTGCACGC GACCTTGCAG CTTTATCGA
 801 AGTGGCTTCC GCACTGGACG CATTTCGCG CGCGGTGAC CAGACCATCG
 851 CCAATCATTG GGTTCCTCCG ACCAGCATCA GCGGCGTAGA ACTGCGTTCC
 901 GCCGTAAACG GCGTGGGTTT CGTTTGGAA GAGACGCGCG GTTCCCATCA
 951 TACGACACAG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCACACAGC
 1001 AGCGTTTAC CAAGCCCTTT TTGGACAACC AGTCCTACAA AGGCTTCAGT
 1051 ATGCTGCTCG ACATCCCGCA CTCTCCGGCA GCGCAAAAAA CTTTCGACGA
 1101 TTTGTTTATG GATTGGCGGG TACGCGCTGC CGGCAAGTTG AACCTGAATC
 1151 TGGTCAACGA CAAAATGGAA GAAGTTTCGA CCCAATGSGT CAAAGACGTG
 1201 CGCACTTATG TATTGGCGCG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA
 1251 ACCGGGCGCG AAAACCGCAT TCGCGCTGTT CTCCTAA

This corresponds to the amino acid sequence <SEQ ID 526; ORF119-1>:

15 1 MIYIVLFLAV VLAVVAYNMY QENQYRKVKR DQFGHSDKDA LLNSKTSHVR
 51 DGKPSGGGSM MPKPQPAVKK TAKPQDPAMR NLQEQDAVYI AKCKQAKAS P
 101 FKTEIETALE ESLIGNSAH TVSEPQTGHS APKPADAPAK PAPVPQTFAK
 151 FLITLKLKSL VELPWFVVR FDTISYIALTE AKELHALPRL SNRCRYQIVG
 201 CTMDHDPQIA EPIGTATQQA FTVGIAQVSR NGLASQELS AFNRQVDAFA
 251 QSMGGQTFLH DLAAFTIEVAS ALDAFCARVD QTIAIHLYSP TSISSVELES
 301 AVTGVGVLE DDGAFHYTDT SGSTMFSSICS LNNPFTNAL LDNQSYKGFS
 351 MLLDIPHSFA GEKTFDDLFM DLAVRLSGQL NLNLVNDPME EVSTQWLKDV
 401 RTYVLARQSE MLKVGIEPFG KTAIRLFS*

Computer analysis of this amino acid sequence gave the following results:

25 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF119 shows 93.7% identity over a 175aa overlap with an ORF (ORF119a) from strain A of *N.*

meningitidis:

		10	20	30	40	50	60
30	orf119.pep	MIYIVLFLAVVLAVVAYNMYQENQYRKVKRDQFGHSDKDAL	LSXTS	SHVR	DGKPSGGGSM		
	orf119a	MIYIVLFLAAVLAVVAYNMYQENQYRKVKRDQFGHSDKDAL	LSXTS	SHVR	DGKPSGGGSM		
		10	20	30	40	50	60
35	orf119.pep	MPKPQPAVKKTAKPQDPXMRNLQEQDAVYIAKQKAKASPF	KTEIETALEES	GIIGNSAH			
	orf119a	MEKPQPAVKKTAKSQDPAMRNLQEQDAVYIAKQKAKASPF	KTEIETALEES	GIIGNSAH			
		70	80	90	100	110	120
40	orf119.pep	TVSEPQTGHSATKPADASAKPAPVPQTFAKPLITLKLKSL	VELSWFDRIDFISY				
	orf119a	TVPEPQTGHSAPKPADAPAKPVPVQTFAKPLITLKLKSL	VELPWFDRVDFISYIALTE				
		130	140	150	160	170	180
45	orf119a	AKELHALPRLSNRCRYQIVGCTMDHDPQIAEPIGRIYQAF	IVIGIQAQVSRNGLASQELS				
		190	200	210	220	230	240

The complete length ORF119a nucleotide sequence <SEQ ID 527> is:

50 1 ATGATTTACA TCGTACTGTT CCGCGCGCC GTCTCGCGG TTCTCGCCTA
 51 CAATATGTAT CAGGAACACC AATACCGCAA AAAAGTGCAG GACCATTTCC
 101 GGCACCTCGA CAAAGATGCC CTGCTCAACA GCAAAACAG CCATGTCCCG
 151 GACGCGAAAC CGTCCGCGCG GCCAGTCATG ATGCGGAAC CCCAACCGCG
 201 GGTCAAAAAA ACGGCAAAAT CCGAAGACCC CGCCATGCGC AACCTGCAAG
 251 AGCAGGATCG CGTCTACATC GCCAAGCAGA AACAGGCGAA AGCCTCCCGC
 301 TTTCAAAACCG AATTCGAAC CGCCTTGGAA GAAAGCGGCA TTATCGGCAA
 351 CTTCGCGCCAC ACGTTTCCCG AACCCCAAAC CGGACATTTC GCACCAAAAC
 401 CTGCGCAGCG CGCGGCAGAA CCGTGTCCCG TTCGCAAAAC GCGGCAAAAC
 451 CGCGTGATTA CGCTCAAAGA GCTGTGGAAG GTGCGAGCTGC CCTGTTTGA
 501 CGTGCCTTTC GACTTCATCT CTTATATCGC GTGACCGGAA GCCAAGAAGC
 551 TGCAACGACT GCGCGCGCTT TCCAAACGCT GCGCGTACCA GATTGTGCGG
 601 TGCAACATGG ACGACCATTT CCAGATTGCG GAACCCATTC CGGGCATCGC

-309-

651 CTATCAGGCA TTTATCGTGG GTATTGAGG AGTCAGCCGC AACGGACTTGG
 701 CCTCGCAGGA AGAACTCTCC GCATTCAACC GCGAGGTGGA TGCATTGCGA
 751 CACAGCATGG GCGGCTCAGC GCTGCACACC GACCTTGCCG CCTTATTGCA
 801 AGTGGCTTCC GCACTGAGAC CATTCGCGCG GCGCGTGCAG CAGACTATCG
 851 CCATCCATTG GGTTCCTCCG ACCAGCATCA GCGCGGTAGA ACTGCGTTC
 901 GCGGTAAACG GCGTGGGTTT CGTITTTGAA GACGACGGCG CGTCCACTA
 951 TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCACACAAG
 1001 AGCGGTTTAC CAATGCCCTT TTGGACACCC AGTCTATAA AGGCTTCACT
 1051 ATGCTGCTCG ACATCCGCGA CTCTCGCGCA GCGGAAAAAA CCTTCGACGA
 1101 TTTGTTTATG GATTGCGCGG TAGCGCTGTC CGGCGAGTGT AACCTGAATC
 1151 TGGTCAACGA CAAAATGGAA GAAGTTTCGA CCCAATGGCT CAAAGACGCG
 1201 CGCACTTATG TATTGCTCGT TCGTCCGAG ATGCTCAAAG TCGGTATCGA
 1251 ACCGGGCGGC AAAACGCGAT TCGGCTGTT CTCTTAA

This encodes a protein having amino acid sequence <SEQ ID 528>:

1 MIYIVLFLAA VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSVHR
 51 DGKESGGFVM MKRFQPAVKK TAKSQDEPAMR NLQEQDAVIL AKQKQAKASP
 101 FKTEIETALE ESGIIGNSAH TVPEPQTGHS AKPADAPAK PVFVETPAK
 151 PLITLKLKSLK VELFWDFVRF DFTSYIALTE AKELHALPRL SNRCRYQIVG
 201 CTMDHDFQIA EPIPGIRYQA FIVGIQAVSR NGLASQBELS AFNRQVDATA
 251 HSMGGTTLHT DLAAFEIVAS ALDAFCARVD OTIAIHLVPS TSISGVELAS
 301 AVTGGVFVLE DDGAFHYTDT SGSTMFSICS LNNPEFTNAL LDNQSYKGSF
 351 MLLDIPHSPA GEKTFDDLFM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV
 401 RTYVLARQSE MLKVGIEPFG KTALELRF*

ORF119a and ORF119-1 show 98.6% identity in 428 aa overlap:

		10	20	30	40	50	60
25	orf119a.pep	MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDAL	LLNSKTSVHRD	GKPSGGPV			
	orf119-1	MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDAL	LLNSKTSVHRD	GKPSGGPV			
		10	20	30	40	50	60
30	orf119a.pep	MPKQPQPAVKKTA	KSQDPAMRNLQEQDAVIAKQKQAKASPF	FKTEIETALEESGIIGNSAH			
	orf119-1	MPKQPQPAVKKTA	KSPADAMRNLQEQDAVIAKQKQAKASPF	FKTEIETALEESGIIGNSAH			
		70	80	90	100	110	120
35	orf119a.pep	TVPEPQTGHSAPKPADAPAKPVFVETPAKPLITLKLKSLKVELFWDFVRF	DFTSYIALTE				
	orf119-1	TVSEPTGHSAPKPADAPAKPAPVETPAKPLITLKLKSKVELFWDFVRF	DFTSYIALTE				
		130	140	150	160	170	180
40	orf119a.pep	AKELHALPRLSNRCRYQIVGCTMDHDFQIAEPIPGIRYQAFIVGIQAVSRNGLASQBELS					
	orf119-1	AKELHALPRLSNRCRYQIVGCTMDHDFQIAEPIPGIRYQAFIVGIQAVSRNGLASQBELS					
		190	200	210	220	230	240
45	orf119a.pep	AFNRQVDATAHSMGGTTLHTDLAAFEIVASALDAFCARVDOTIAIHLVPS	TSISGVELAS				
	orf119-1	AFNRQVDATAHSMGGTTLHTDLAAFEIVASALDAFCARVDOTIAIHLVPS	TSISGVELAS				
		250	260	270	280	290	300
50	orf119a.pep	AVTGGVFVLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGSF	MMLDIPHSPA				
	orf119-1	AVTGGVFVLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGSF	MMLDIPHSPA				
		310	320	330	340	350	360
55	orf119a.pep	GEKTFDDLFM	DLAVRLSGQLNLNLVNDKMEEVSTQWLKDV	RTYVLARQSEMLKVGIEPFG			
	orf119-1	GEKTFDDLFM	DLAVRLSGQLNLNLVNDKMEEVSTQWLKDV	RTYVLARQSEMLKVGIEPFG			
		370	380	390	400	410	420
60	orf119a.pep	GEKTFDDLFM	DLAVRLSGQLNLNLVNDKMEEVSTQWLKDV	RTYVLARQSEMLKVGIEPFG			
	orf119-1	GEKTFDDLFM	DLAVRLSGQLNLNLVNDKMEEVSTQWLKDV	RTYVLARQSEMLKVGIEPFG			
		370	380	390	400	410	420

orf119a.pep KTAALRLFSX
 |||||
 orf119-1 KTAALRLFSX

5 Homology with a predicted ORF from *N.gonorrhoeae*

ORF119 shows 93.1% identity over a 175aa overlap with a predicted ORF (ORF119ng) from *N.gonorrhoeae*:

orf119.pep	MIYIVLFLAVLAVVAYNMYQENQYRKVRDQFGHSDKDALNLSKTSHVDRGKPSGGSPVM	60
orf119ng	MIYIVLFLAVLAVVAYNMYQENQYRKVRDQFGHSDKDALNLSKTSHVDRGKPSGGSPVM	60
orf119.pep	MPKPQPAVKKTAKPQDPXMRNLQEQDAVYIAKQKQAKASFTEIETALEESGIIGNSAH	120
orf119ng	MPKPQPAVKKTAKPQDSAMRNLQEQDAVYIAKQKQAKASFTEIETALEESGIIGNSAH	120
orf119.pep	TVSEPTQGHSAKPADASAKPAVPVQTPAKPLITLKLKSKVLEWDFVDRIDFISY	175
orf119ng	TVSEPTQGHSAKPADAPAKFVFPVQTPAKPLITLKLKSKVLEWDFVDRIDFISYIALTE	180

The complete length ORF119ng nucleotide sequence <SEQ ID 529> is:

20	1	ATGATTAC	TCGACTGT	CCTCGCGCC	GTCCTCGCG	TTGTCGCCA
	51	CAATATGAT	CAGGAAACC	AATACGCAA	AAAAGTGGC	GACCGATTG
	101	GACACTCGA	CAAGATGCC	CTGCTCAAC	GCAAAACCG	CCATGTCGC
	151	GACGGCAAC	CGTCCGGCG	GCCAGTCAT	ATGCCGAAAC	CCCAACCGC
	201	GGTCAAAAA	CGCGCCAAAC	CCAAGACTC	CGCCATGCG	AACTCGCAAG
25	251	AACAGGATGC	CGTCTACATC	GCAAGCAGA	AACAGGCAAA	AGCCTCCCG
	301	TTCAAAACCG	AAATCGAAAC	CGCCTTGAA	GAATCGGCA	TTATCGGCA
	351	CTCCGCCAC	ACCGTTTCG	AACCCCAAC	CGGACATTCC	GCACGAAAC
	401	CTGCCGACG	CGCGGCAAAA	CCCGTTCCG	TTCCGCAAA	CGCCGCAAA
	451	CGCGTGATTA	CGCTCAAGA	GCTGTGCAAG	GTCGAGCTG	CTCGGTTGA
30	501	CGTGGCTTC	gACTGACTCT	CCTATATCG	GCTGACGAA	GCCAAAGAC
	551	TGCACGACT	GCGTCCGCTT	CGCGCTACA	GCGCTTACG	CGGCTTACG
	601	TCACCATGG	ACGACCATTT	CCAGATTCG	GACACATCC	CGGCTTACG
	651	CTATCAGCA	TTTATCGTG	GTATCAGGC	AGTCAGCGC	AACGACTTG
	701	CTCGCAGGA	AGAACCTCC	GCATTCAAC	GCCAGGCGA	CGCATTGCA
35	751	CAAGCATGG	GCGGTCAGC	GCTGCACAC	GACCTTGCG	CCTTATCGA
	801	AGTGGCTTC	GCACTGGAC	CATTCTGCG	GCGGTCGAC	CAGACCATG
	851	CCATCCATTT	GGTTTGGCG	ACCAGCATC	GCGGCTAGA	ACTCGGTTCC
	901	CGCGTACGG	CGGTGGGTT	CGTTTGGAA	GACGACGGC	CGTTCCTCA
	951	TACCGCACG	TCGGGCTCG	CCATGTTCT	CATCTGCTG	CTCAACAAG
40	1001	AGCGGTTTAC	CAATGCCCTT	TTGGACAAC	AGTCTCTACA	AGGCTTCAGT
	1051	ATGCTGCTCG	ACATCCCGCA	CTCTCCGCC	GCGCAAAAA	CCTTCAGCA
	1101	TTTGTATTATG	GATTTGGCG	TACGCTGTC	CGGTGAGTT	AACCTGAATC
	1151	TGGTCAACGA	CAAAATGGAA	GAGATTTCGA	CCCAATGGCT	CAAGACGTA
	1201	CGCATTATG	TATTTGGCG	TGACTCGAG	ATGCTCAAAG	TCGGTATCGA
45	1251	ACCGGGCGGC	AAAACGCCC	TGCGCTGTT	TTCTATA	

This encodes a protein having amino acid sequence <SEQ ID 530>:

1	MIYIVLFLAA	VLAVVAYNMY	QENQYRKVR	DQFGHSDKA	LLNSKTSVHR
51	DGKPSGGPVM	MPKPQPAVK	PAKQDSAMR	NLQEQDAVY	AKQKQAKASP
101	PKTEIETALE	ETIGIIGNSAH	TVSEPTQGH	APKPADAPK	FVFPVQTPAK
151	PLITLKLSEK	VELFWDFVRF	DEISYIALTE	AKELHALPEL	SNRCRYQIVG
201	CTMDHFIQIA	EPIPCIRYQA	FVGTQAVSR	NGLASQELSL	AFNRQADAF
251	QSMGGQT LHT	DLAFLIEVAS	ALDAFCARVD	QTIAIHLVSP	TSISGVLESL
301	AVTGVGFVLE	DDGAFHYTDT	SGSTMFSICS	LNNEPFTNAL	LDNQSYKGFS
351	MLLDIPHSFA	GEKTFDDLFM	DLAVRLSGQL	NLNLVNDKME	EVSTQWLKDV
55	401	RTYVLARQSE	MLKVGIEPGG	KTAALRLFS*	

ORF119ng and ORF119-1 show 98.4% identity over 428 aa overlap:

	10	20	30	40	50	60
orf119ng	MIYIVLFLAAVLAVVAYNMYQENQYRKVRDQFGHSDKDALNLSKTSHVDRGKPSGGSPVM					
orf119-1	MIYIVLFLAVLAVVAYNMYQENQYRKVRDQFGHSDKDALNLSKTSHVDRGKPSGGSPVM					
	10	20	30	40	50	60

		70	80	90	100	110	120
	orf119ng	MPKPPQPAVKKPAKPPQDSAMRNLEQEDAVYIAKQKQAKASF	FKTEIETALEE	IGIIGNSAH			
5	orf119-1	MPKPPQPAVKKPAKPPQDPAMRNLEQEDAVYIAKQKQAKASF	FKTEIETALEE	SGIIGNSAH			
		70	80	90	100	110	120
	orf119ng	TVSEPTGHSAPKPADAPAKPVVPTPAKPLITL	KELSKVLPWF	DVRFDFISYIALTE			
10	orf119-1	TVSEPTGHSAPKPADAPAKPVVPTPAKPLITL	KELSKVLPWF	DVRFDFISYIALTE			
		130	140	150	160	170	180
	orf119ng	AKELHALFRLSNRCRYIVGCTMDHFPQIAEP	IPGIRYQAFIVGIQAVSRNGLASQ	BEELS			
15	orf119-1	AKELHALFRLSNRCRYIVGCTMDHFPQIAEP	IPGIRYQAFIVGIQAVSRNGLASQ	BEELS			
		190	200	210	220	230	240
	orf119ng	AFNRQDAFAQSMGGQTLHTDLAAFIEVASALDAFCARV	DOTIAIHLVSP	TPSISGVELRS			
20	orf119-1	AFNRQDAFAQSMGGQTLHTDLAAFIEVASALDAFCARV	DOTIAIHLVSP	TPSISGVELRS			
		250	260	270	280	290	300
	orf119ng	AVTGVGFLVLEDDGAFHYTDSGSTMFSICSLNNEP	FTNALLDNQSKGFSMLLDIPHSPA				
25	orf119-1	AVTGVGFLVLEDDGAFHYTDSGSTMFSICSLNNEP	FTNALLDNQSKGFSMLLDIPHSPA				
		310	320	330	340	350	360
	orf119ng	GEKTFDDLFLMDLAVRLSGQLNINLVNDRKMEEVSTQ	WLKDVRTYVYLARQSEMLKVGIEP	GG			
30	orf119-1	GEKTFDDLFLMDLAVRLSGQLNINLVNDRKMEEVSTQ	WLKDVRTYVYLARQSEMLKVGIEP	GG			
		370	380	390	400	410	420
	orf119ng	KTALRLFSX					
35	orf119-1	KTALRLFSX					
		429					
40	orf119ng	KTALRLFSX					
	orf119-1	KTALRLFSX					

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

45 Example 64

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 531>

	1	..GCGCGGCACG	GCA	CGGAAGA	TTTCTTCATG	AACACACGCG	ACAC.ATCAG
	51	GCAGATAGTC	GA	ANGCACC	CCGGTACGAT	GAAGCTGCTG	ATTTCCTCCA
50	101	TCGCCCTGAT	TT	CATTGGTA	GTCGCGCGCA	TCGGCGTGAT	GAACATCATG
	151	CTGGTGTCG	TT	ACCGAGCG	CACCAAAGAA	ATCGGCATAC	GGATGCCAAT
	201	CGGCGCGCGG	CG	CGGCAATA	TTTTCGAGCA	GTTTTTGATT	GAGGCGGTGT
	251	TAATCTGCGT	CAT	CGGCGCGT	TTGGTCGCGG	TGGGTTTGTC	CGCGCGCGTC
	301	AGCGCTGCT	TC	AATCATTT	TGTAACCCAC	TTCCCGATGG	ACATTTCCGC
	351	CATGCTGCTC	AT	CGGCGCGG	TCGCGCTGTC	GACCGGAATC	GGCATCGCGT
55	401	TCGGCTTTAT	GC	CTGCCAAT	AAAGCAGCCA	AACATCAATC	GATAGACGCA
	451	TTGGCAGCAG	ATT	GAA			

This corresponds to the amino acid sequence <SEQ ID 532; ORF134>:

	1	..ARHGTEDDFM	NNSDXIRQIV	ESTTGTMKLL	ISSIALISLV	VGGIGVMNIN
	51	LVSVTERTEKE	IGIRMAIGAR	RGNIXQOFLI	EAVLVICVIG	LVGVLGSAAV
60	101	SLVNFHFVTD	FPMDISAMSV	IGAVACSTGI	GIAFGFMPAN	KAAKLNPIDA
	151	LAQD*				

Further work revealed the complete nucleotide sequence <SEQ ID 533>:

```

1   ATGTCGGTGC AAGCAGTATT GGGGCACAAA ATGCGTTCGC TTCTGACGAT
51  GCTCGGCATC ATCATCGGTA TCGCGTCGGT GGTTCCTGTC GTGCGATTGG
101 GCAATGGTTC GCAGAAAAAA ATCCTTGAAG ACATCAGTTC GATAGGGACG
151 AACACCATCA GCATCTTCCC GGGGCGCGCG TTGCGGGACA GGGCAGCGCG
201 CAGGATTAAA ACCCTGACCA TAGACGACGC AAAAATCATC GCCAAACAAA
251 GCTACGTTGC TTCCGCGCAG CCATGACTTT CGAGCGGCGG CAGCGTGAAT
301 TACCGCACCA CGACCTTGAC CCGCTCGCTT TACGGGTGGG GCGAACAAAT
351 TTTTCGACCTG CGGCGACTGA ACGTGCAGAA GGGGCGGCTG TTTGACGAAA
401 ACGATGTGAA AGAAGACGCG CAGGTGCTCG TCATCGACCA AAATGTCAAA
451 GACAAACTCT TTGCGACTC GATCCGTTG GGTAAAAACA TTTTGTTCAG
501 GAAACGCCCC TTGACCGGTA TCGCGTGTAT GAAAAAGAC GAAACGCTT
551 TCGGCAATTC CGACGTGCTG ATGCTTTGGT CGCCCTATAC GACGTGATG
601 CACCAATATC CAGGCGAGAG CCACACCAAC TCCATCACOG TCAAAATCAA
651 AGACAATGCC AATACCCAGG TTGCGGAAAA AGGGCTGACG GATCTGCTCA
701 AAGCGCGGCA CGCACCGGAA GATTTCCTTA TGAACAACAG CGACGACATC
751 AGGCAGATAG TCGAAAGCAC CACCGGTACG ATGAAGCTGC TGATTTCCTC
801 CATCGCCCTG ATTTCAATTG TAGTCGCGCG CATCGCGGTG ATGAACATCA
851 TGCTGGTGTG CGTTACCGAG CGCACCAAG AAATCGGCTA ACGGATGGCA
901 ATGCGCGCGC GGGCGGCGAA TATTTTTCAG CAGTTTTTGA TTGAGCGGCT
951 GTTAATCTGC GTCACTGGCG GTTTGTCGCG CGTGCGTTTG TCGCGCGCGC
1001 TCAGCTCGT GTTCAATCAT TTTGTAACCG ACTTCGCGAT GGACATTTCC
1051 GCGATGTCG TCATCGGCGC GGTGCGCTGT TCGACCGGAA TCGGCATCGC
1101 GTTCGCGCTT ATGCTGCCA ATAAAGCAGC CAAACTCAAT CCGATAGACG
1151 CATTGGCACA GGATTGA

```

This corresponds to the amino acid sequence <SEQ ID 534; ORF134-1>:

```

1   MSVQAVLAHK MRLSLTMLGI IIGIASVVSU VALNGSQKK ILEDISSIGT
51  NTISIFPGRG FGDRRSGRK TLTIDDAKII AKQSYVASAT PMTSGGTLT
101 YRNTDLTASL YGVGEQYFDV RGLKLETGRL FDENVDKEDA QVVVIDQNVK
151 DKLFADSDPL GKTILFRKRP LTVIGVMKKD ENAFGNSDVL MLMSPTTVM
201 HQITGESHTN SITVKIKDNA NTQVAEKLGT DLLKARHGT EDFMNNSDSI
251 RQIVESTTGT MKLLISSIAL ISLVVGGIGV MNIMLVSVTE RTKEIGIRMA
301 IGARRGNILQ QFLIEAVLIC VIGGLVGVGL SAAVSLVFNH FVTDFDFMDSI
351 AMSVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical protein o648 of *E. coli* (accession number AE000189)

ORF134 and o648 protein show 45% aa identity in 153aa overlap:

```

Orf134: 2   RHGTEDFFMNNSDXIRQIVESTTGTMKXXXXXXXXXXXXXVVGIGVMNIMLVSVTERTKEI 61
              RHG +DFF N D + + VE TT T++                VVGIGVMNIMLVSVTERT+EI
40  o648: 496 RHGKDFFTWNMDGLTKVEKTRTTLQLFLTLVAVISLVVGIGVMNIMLVSVTERTREI 555
Orf134: 62   GIRMAIGARRGNIXQFLIEAXXXXXXXXXXXXXXXXXXFNHFVTFDFPMDISAMSVI 121
              GIRMA+GAR ++ QFLIEA                        F+ + + S ++++
45  o648: 556 GIRMVAGARASDLVQFLIEAVLVCLVGGALGITLSLLIAFTLQLFLFEGWEIGFSLPLL 615
Orf134: 122  GAVACSTGIGIAFGFMANKAAKLNPIDALAQD 154
              A CST GI FG++PA AA+L+P+DALA++
o648: 616 LAFLCSTVTGILFGWLPARNAARLDVVDALARE 648

```

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF134 shows 98.7% identity over a 154aa overlap with an ORF (ORF134a) from strain A of *N. meningitidis*:

```

10      20      30
55  orf134.pep  ARHGTEDFFMNNSDXIRQIVESTTGTMKLL
              |||||
orf134a  GESHTNSITVKIKDNANTQVAEKLGTDLLKARHGTEDFFMNNSDXIRQIVESTTGTMKLL
              210      220      230      240      250      260
              40      50      60      70      80      90

```

orf134.pep	ISSIALISLVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNIXQQFLIEAVLICVIGG	
orf134a	ISSIALISLVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNILQQFLIEAVLICVIGG	
	270 280 290 300 310 320	
orf134.pep	LVGVGLSAAVSLVFNHFVTFDPMDISAMSVIGAVACSTGIGIAFGMPANKAARKLNPIDA	100 110 120 130 140 150
orf134a	LVGVGLSAAVSLVFNHFVTFDPMDISAMSVIGAVACSTGIGIAFGMPANKAARKLNPIDA	
	330 340 350 360 370 380	
orf134.pep	LAQDX	
orf134a	LAQDX	

The complete length ORF134a nucleotide sequence <SEQ ID 535> is:

1	ATGTCGGTGC	AAGCACTATT	GGCGCACAAA	ATGCGTTCGC	TTCTGACGAT
51	GCTCGGCATC	ATCATCGGTA	TCGCTTCGGT	TGCTCCGCT	GTCGCATTGG
101	GCACGCGTTC	GCAGAAAAAA	ATCCTTGAAG	ACATCACTTC	GATAGGAGCG
151	AACACCATCA	GCATCTTCCC	AGGGCGCGGC	TCGCGGACA	GGCGCAGCGG
201	CAGGATTAA	ACCTCGACCA	TAGACGACGC	AAAATCATC	GCCAAACAA
251	GCTACGTTGC	TTCCGCCACG	CCCATGACTT	CGAGCGCGCG	CACGCTGACT
301	TACCGCAATA	CGACCTGAC	CGCTTCTTTG	TACGGTGTGG	GCGAACAATA
351	TTTCGACGTG	CGCGGGCTGA	AGCTGGAAAC	GGGGCGGCGT	TTTGACGAAA
401	ACGATGTGAA	AGAAGACGCG	CAGGTGCTCG	TCATCGACCA	AAATGTCAAA
451	GACAACTCT	TTGCGGACTC	GGATCGGTG	GGTAAACCA	TTTTGTCGAC
501	GAACGCCGCC	TTGACCGTCA	TCGGCGTGAT	GAAAAAGAC	GAAGACGCTT
551	TCGGCAATTC	CGACGTGCTG	ATGCTTTGGT	CGCCCTATAC	GACGGTGATG
601	CACCAATCA	CAGCGAGAG	CCACACCAAC	TCCATCACCG	TCAAAATCAA
651	AGACAATGCC	AATACCCAGG	TTGCCGAAA	AGGGCTGACC	GATCTGCTCA
701	AAGCGCGCGA	CGCACGGGAA	GATTTCCTCA	TGAACAACAG	CGACAGCATC
751	AGGCAGATAG	TCGAAGACAC	CACCGTAGCG	ATGAAGCTCG	TGATTTCTCT
801	CATCGCCCTG	ATTTCATTGG	TAGTCGGCGG	CATCGCGCGT	ATGACATCAT
851	TCGCGGTGTC	CGTACCCGAG	CGCACCAAG	AAATCGGCAT	ACGATGCGCA
901	ATCGCGCGCG	GGCGGGCGAA	TATTTTCGAC	CAGTTTGTGA	TTGAGCGCGT
951	GTTAATCTGC	GTCATCGCGG	GTTTGCTCGG	CGTGCGGTTG	TCGCGCGCGG
1001	TCAGGCTCGT	GTTCAATCAT	TTTGTAAACG	ACTTCCGAT	GGAATATTCG
1051	GCCATGTCG	TCATCGCGCG	GGTCGCTGT	TCGACCGGAA	TCGCGATCCG
1101	GTTCCGGCTT	ATGCTGCGCA	ATAAAGCAGC	CRAACTCAAT	COGATAGATG
1151	CATTGGCGCA	GGATTGA			

This encodes a protein having amino acid sequence <SEQ ID 536>:

1	MSVQAVLAHK	MRSLLTMLGI	IIGIASVVS	VALNGSQKK	ILEDISSIGT
51	NTISIFPGRG	FGDRRSGRK	TLTIDDAKII	AKQSVASAT	PMTSSGGTLT
101	YRNTDLTASL	YGVGGEQFDV	RGLKLETGR	FDENVKEDA	QVVVDQNVK
151	DKLFADSDPL	KTLLFRKR	LTIVGMKKD	ENAFGNSDVL	MLWSPYTTVM
201	HQITGESHTN	SITVKIKDNA	NTQVAEKGLT	DLKARHGTE	DFMNNSDSI
251	RQIVESTTGT	MKLLISSIAL	ISLVVGGIGV	MNMLVSVTE	RTKEIGIRMA
301	IGARRGNILQ	QFLIEAVLIC	VIGGLVGVGL	SAARSLVFNH	FVTFDFMDIS
351	AMSVIGAVAC	STGIGIAFG	MPANKAARKLN	PIDALAQD*	
50	ORF134a and ORF134-1 show 100.0% identity in 388 aa overlap:				
orf134a.pep	MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALNGSQKKILEDISSIGTNTISIFPGRG				
orf134-1	MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALNGSQKKILEDISSIGTNTISIFPGRG				
orf134a.pep	FGDRRSGRKTLTIDDAKIIAKQSVASATPMTSSGGTLTYRNTDLTASLYGVGGEQFDV				
orf134-1	FGDRRSGRKTLTIDDAKIIAKQSVASATPMTSSGGTLTYRNTDLTASLYGVGGEQFDV				
orf134a.pep	RGLKLETGRLEFENDVKEDAQVVVDQNVKDKLFADSDPLGKTLFRKRPLTVIGVMKKD				
orf134-1	RGLKLETGRLEFENDVKEDAQVVVDQNVKDKLFADSDPLGKTLFRKRPLTVIGVMKKD				
orf134a.pep	ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLKARHGTE				
orf134-1	ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLKARHGTE				

5	orf134a.pep	DFFMNNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLSVSTERTKEIGIRMA
	orf134-1	DFFMNNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLSVSTERTKEIGIRMA
10	orf134a.pep	IGARRGNILQOFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTFDFMDISAMSVIGAVAC
	orf134-1	IGARRGNILQOFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTFDFMDISAMSVIGAVAC
10	orf134a.pep	STGIGIAFGFMPANKAAKLNPIDALAQDX
	orf134-1	STGIGIAFGFMPANKAAKLNPIDALAQDX

Homology with a predicted ORF from *N.gonorrhoeae*

- 15 ORF134 shows 96.8% identity over a 154aa overlap with a predicted ORF (ORF134.ng) from *N. gonorrhoeae*:

20	orf134.pep	ARHGTEDFMNSDXIRQIVESTTGTMKLL 	30
	orf134.ng	GESHNTSITVKIKDNANTRVAEGLAELLKARHGTEDFMNSDSIRQMVESTTGTMKLL 	264
20	orf134.pep	ISSIALISLVVGGIGVMNIMLSVSTERTKEIGIRMAIGARRGNILQOFLIEAVLICVIGG 	90
	orf134.ng	ISSIALISLVVGGIGVMNIMLSVSTERTKEIGIRMAIGARRGNILQOFLIEAVLICVIGG 	324
25	orf134.pep	LVGVGLSAAVSLVFNHFVTFDFMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA 	150
	orf134.ng	LVGVGLSAAVSLVFNHFVTFDFMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA 	384
30	orf134.pep	LAQD 154 	
	orf134.ng	LAQD 388	

The complete length ORF134ng nucleotide sequence <SEQ ID 537> is:

35	1	ATGTGCGTGC	AAGCAGTATT	GGCGCAAAA	ATGCGTTCCG	TTCTGACCAT
	51	GCTCGGCATC	ATCATCGGTA	TCGCTTCGGT	TGTCTCCGTC	GTCGCGCTGT
40	101	GCAACGGTTC	GCAGAAAAAA	ATCTCCGAAG	ACATCAGGTC	GATGGGGACG
	151	AACACCATCA	GCATCTTCCC	CGGGCGGGC	TTCCGCGACA	GGCGACGCGG
45	201	CAAAATCAAA	ACCCTGACCA	TAGACGACGC	AAAATCATC	GCCAAACAAA
	251	GCTACGTTGC	CTCCGCCACG	CCCATGACTT	CGAGCGCGGG	CACGCTGACC
50	301	TACCGCAATA	CCGACCTGAC	CGCTTCTTTG	TACGCTGTGG	GCGAACATA
	351	TTTCGACGTG	CGCGGGCTGA	AGCTGGAAAC	GGGCGCGCTG	TTTGATGAGA
55	401	ACGATGTGAA	AGAAGACGCG	CAAGTCGTCG	TATCATGACCA	AAATGTCAAA
	451	GACAAACTCT	TTGCGGACTC	GGATCCGTTG	GGTAAACACA	TTTTGTTGAG
60	501	GAAACGCCCC	TTGACCGCTA	TCGCGCTGAT	GAAAAAAGAC	GAAAAAGCTT
	551	TCGCGCAATC	CGACCTGCTG	ATGCTTTGCT	CGCCCTATAC	GACCGGTGAT
65	601	CACCAATATC	CMGGCGAGAG	CCACACCAAC	TCCATCACCG	TCAAAATCAA
	651	AGACAATGCC	ANTACCCTGG	TTGCGGAAAA	AGGCTGGCG	GAGCTGCTCA
70	701	AAGCACGGCA	CGGCACGGAA	GACTTCTTTA	TGAACAACAG	CGACGACATC
	751	AGGCAGATGG	TCGAAAGCAC	CACCGGTACG	ATGAAGCTGC	TGATTTCTCT
75	801	CATCGCCCTG	ATTTCATTGG	TAGTCGGGCG	CATCGGTGTG	ATGAACATTA
	851	TGCTGGTGTC	CGTTACCGAG	CGCACAAAG	AAATCGGCAT	ACGGATGGCA
80	901	ATCGGCGCGC	GGCGCGGCAA	TATTTTGCAG	CAGTTTTTGA	TTGAGCGCGT
	951	GTTAATCTGC	ATCATCGGAG	GCTTGGTGCG	CGTAGTTTTC	TCCCGCCCGC
85	1001	TCAGCGCTCT	GTTCAATCAT	TTTGTAACCG	ATTTCCGAT	GGACATTTTC
	1051	CGCGCATCCG	TTATCGGGGC	GCTCGCCTGT	TCGACCGGAA	TCGCGATCGC
90	1101	GTTCCGCTTT	ATGCTCGCCA	ATAAGGCAGC	CAAACTCAAT	CCGATAGATG
	1151	CATTGGCGCA	GGATTGA			

This encodes a protein having amino acid sequence <SEQ ID 538>:

60	1	NSVQAVLAHK	MRSLLTMIGI	IIGIASVVSV	VALNGNSQKK	ILEDISSMGT
	51	NTISIFPFRG	FGDRRSQKIK	TLTIDDAKII	AQOSVVASAT	PMTSSGGTLT
65	101	YRNTDLTASL	YGVGEQYFDW	RLGKLETGR	FDENDVKEDA	QVVDIDONVK
	151	DKLPDSDPL	GKTLIFRKRF	LTIVGVMKGD	ENAFRNSDVL	MLWSPTTTVM
70	201	HQVTCESITN	SITVKIKDNR	MTVNLKGLA	ELLKRNKTE	DFMNSDSIR
	251	RQMVESITGT	MLTSSIAL	ISLVVGGIGV	NIMLSVSTE	RTKEIGIRMA
75	301	IGARRGNILQ	FLIEAVLIC	IIGGLVGVGL	SAAVSLVFNH	FVTFDFMDIS

351 AASVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD*

ORF134ng and ORF134-1 show 97.9% identity in 388 aa overlap:

```

5  orf134ng      MSVQAVLAHKMRSLLTMLGIIIGIASVSVSVLGNQSGKKILEDISSMGTNTISIFPGRG
   orf134-1     MSVQAVLAHKMRSLLTMLGIIIGIASVSVSVLGNQSGKKILEDISSIGTNTISIFPGRG

10 orf134ng      FGDRRSRGKIKTLTIDDAKIIAQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFV
   orf134-1     FGDRRSRGKIKTLTIDDAKIIAQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFV

15 orf134ng      RGLKLETGRFLDENDVKEDAQVVVIDQNVKDKLFADSDPLGKLTILFRKRPLTVIGVMKKD
   orf134-1     RGLKLETGRFLDENDVKEDAQVVVIDQNVKDKLFADSDPLGKLTILFRKRPLTVIGVMKKD

20 orf134ng      ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTRVAEKGAEALLKARHGT
   orf134-1     ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEGLTDLKARHGT

25 orf134ng      DFFMNNSDSIRQMVESTTGTMKLLISSIALISLVGGIGVMNIMLVSVTERTEIGIRMA
   orf134-1     DFFMNNSDSIRQIVESTTGTMKLLISSIALISLVGGIGVMNIMLVSVTERTEIGIRMA

30 orf134ng      IGARRGNILQOFLIEAVLICIGLVGGLSAAVSLVFNHFVTFPMDISAASVIGAVAC
   orf134-1     IGARRGNILQOFLIEAVLICIGLVGGLSAAVSLVFNHFVTFPMDISAMSIGAVAC

   orf134ng      STGIGIAFGFMPANKAAKLNPIDALAQDX
   orf134-1     STGIGIAFGFMPANKAAKLNPIDALAQDX

```

30 ORF134ng also shows homology to an *E. coli* ABC transporter:

```

sp|P75831|YBJZ_ECOLI_HYPOTHETICAL_ABC_TRANSPORTER_ATP-BINDING_PROTEIN_YBJZ_gi5
(AE000189) o648; similar to YBBA_HAEIN SW: P45247 [Escherichia coli] Length =
648
35  Score = 297 bits (753), Expect = 6e-80
   Identities = 162/389 (41%), Positives = 230/389 (58%), Gaps = 1/389 (0%)

Query: 1  MSVQAVLAHKMRSLLTMLXXXXXXXXXXXXXGNQSGKKILEDISSMGTNTISIFPGRG 60
      M+ +A+ A+KMR+LLTML +G+ +++ +L DI S+GNTPI ++PG+
40  Sbjct: 260 MAMRALAANKMRTLLTMLGIIIGIASVSVSVVGDAAKQWVLADIRSIGNTIDVYPGKD 319

Query: 61  FGDRRSRGKIKTLTIDDAKIIAQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFV 120
      FGD + L DD I KQ +VASATP S L Y N D+ AS GV YF+V
Sbjct: 320 FGDDDPYQQAQKLYDDLIAIQKQPWVASATPAVSQNLRLRYNNVDVAASANGVSGDYFNV 379

45  Query: 121 RGLKLETGRFLDENDVKEDAQVVVIDQNVKDKLFADSDPLGKLTILFRKRPLTVIGVMKK 179
      G+ G F++ + AQVVVD N + +LF +D +G+ IL P VIGV ++
Sbjct: 380 YGMTFSGNTFNQEQNLGRAQVVVDLSDNTRQLFLPHKADVVGEVILVGNMPEARVIGVAEE 439

50  Query: 180 DENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTRVAEKGAEALLKARHGT 239
      ++ FG+S VL +W FY+T+ ++ G+S NSITV++K+ ++ AE+ L LL RHG
Sbjct: 440 KQSMFGSSKVLKVLWLPYSTMSGRVMGQSWLNS ITVRVKEGFDSEAEQQLTRLLSLRHGK 499

55  Query: 240 EDDFMNNSDSIRQMVESTTGTMKKXXXXXXXXXXXXXVGGIGVMNIMLVSVTERTEIGIRMA
      +DFF N D + + VE TT T++ VVGGIGVMNIMLVSVTERTEIGIRM
Sbjct: 500 KDDFTWNMDGVLKTVKTRTTLQLFLTLVAVISLVVGGIGVMNIMLVSVTERTEIGIRM 559

Query: 300 AIGARRGNILQOFLIEAVLICIGLVGGLGILTSLLIAFTLQLFLPGWEIGFSLALLLAFL 359
      A+GAR + +LQOFLIE +F+ + + S +++ A
Sbjct: 560 AVGARASDVLQOFLIEAVLICVGLGGLGILTSLLIAFTLQLFLPGWEIGFSLALLLAFL 619

60  Query: 360 CSTGIGIAFGFMPANKAAKLNPIDALAQD 388
      CST GI FG++PA AA++L+P+DALA++
Sbjct: 620 CSTVTGILFGWLPAARNAARLDPVDALARE 648

```

Based on this analysis, including the presence of the leader peptide and transmembrane regions in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 65

- 5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 539>:

```

1   ..GGGACGGGAG CGATGCTGCT GCTGTTTAC GGGTAACGA T.CTGCCTTT
51  GGCCACTGGC GTTACCCGTA GTTACACCTC GTCGATTTTT TTGGCGGTAT
101 TTTCCTTCCT GATTTTGAAA GAACGGATTT CGGTTTACAC GCAGGCGGTG
151 CTGCTCCTTG GTTTTGCCGG CGTGGTATTG CTGCTTAAC CCTCGTTCGG
201 ACGCGCTCAG GAACCGCGCG CACTCCCGCG GCTGCGCGGG GGGCGAATGT
251 CGCGCTGGCG GTATTTGAAA GTGCGGGAAC TGTCTTTGGC GGGCGAACCT
301 GGCTGGCGCG TCGTGTTTTA CCTTCCGTCG ACAGGTGTGG CGATGTCGCT
351 GGTTTGGGCG ACCTGACGCG CTTGGCACAC CTGTCCTTT CCATCGGCAG
401 TTTATCTGTC GTGCATCGCG GTGTCGCGCG TGATTGCCCA ACTGTCGATG
15 451 ACGCGCGCCT ACAAAGTCGG GCACAAATTC ACGGTTGCTT CGCTTTTCTTA
501 TATGACCGTC GTTTTTTCCG CTCTGTCTGC CGCATTTTT CTGGCGGAAG
551 AGCTTTTCTG GCAGGAAATA CTCGGTATGT GCATCATCAT CCTCAGCGGT
601 ATTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 540; ORF135>:

```

20 1   ..GTGAMLLFY AVTILPLATG VTLSTYSSIF LAVFSFLIK ERISVYTQAV
51  LLLGFAGVVL LLNPSFRSGQ ETAALAGLAG GMSGWAYLK VRELSLAGEP
101 GWRVVFYLSV TGVAMSSVMA TLTGWHITLSF PSAYVLSLIG VSALIAQLSM
151 TRAYKVGDKF TVASLSYMTV VFSALSAFFL LGEELEWQEI LGMCIISAV
201 F*

```

- 25 Further work revealed the complete nucleotide sequence <SEQ ID 541>:

```

1   ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGCG
51  GGGGCGCTGC TTTACATTAA TGAACGATTT GATTAAAGAG GCATCGGCAG
101 AATTTGCCCT CGGCAGCGCG GAATGAGTCT TTTGCGCGAT GCTGTTTTCA
151 ACCTGTGCGC TCGGCGCTGC CGCGGTATTG GCTCGGAGCA MCTTCCGAC
30 201 GCGCCATTGG AAAAACCATC TAAACCGCAG TATGCTCGGG ACGGGGCGGA
251 TGCTGTGCTC GTTTTACGCG GTAACGATC TGCCCTTTGGC CACTGGCGGT
301 ACCCTGAGTT ACACTCGTGC GATTTTGTG CGGCTATTTT CTTCTGTGAT
351 TTTGAAAGAA CGGATTTCCG TTTACACGCA GCGGTGCTG CTCTCTGGTT
401 TTGCGCGCGT GGTATTGCTG CTTAATCCCT CGTTCCGAGC CGGTACAGAA
35 451 ACGCGCGCAC TCGCGCGGCT GCGCGCGCGC GCGATGTCGG GCTGGCGGTA
501 TTTGAAAGTG CGCGAAGTGT CTTTGGCGGG GCACCCGCGC TGGCGCGTCG
551 TGTTTTACCT TTCCGTGACA GGTGTGCGCA TGTCTCGGTT TTGGGCGAGC
601 CTGACCGCGT GGCACACCTT GTCCTTTCCA TCGGCAGTTT ATCTGTCGTG
651 CATCGCGCTG TCGCGCTGTA TTGCCCAACT GTCGATGAGC CGCGCTACCA
40 701 AAGTCGCGCA CAATTTCAAG GTTGCGCTGC TTTCCTATAT GACCGCTGTT
751 TTTTCCGCTC TGCTGCGCGC ATTTTTCCTG GCGCAAGAGC TTTTCTGGCA
801 GGAAATACTC GGTATGTGCA TCATCATOCT CAGCGGTATT TTGAGCAGCA
851 TCGCCCCAC TGCTTCAAAA CAGCGCGTGC AATCCCTGTT CGCGCAAGAA
901 TAA

```

- 45 This corresponds to the amino acid sequence <SEQ ID 542; ORF135-1>:

```

1   MDTAKKDILG SGWMLVAAC FTIMNVLIKE ASAKFALGSG ELVFRMFLFS
51  TVALGAARVL RDXFRTPHW KHLNRSWVG TCGAMLLFYA VTHLPLATGV
101 TLYSTSIIFL AVFSFLIKE RLSVYTOAVL LLGFAGVVL LLNPSFRSGQ
151 TAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSVT GVMSSVKAT
201 LTGWHITLSF SAVYLSLIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV
50 251 FSALSAFFL LGEELEWQEI LMCIIISGI LLSIRPTAFK QRLQSLFRQR
301 *

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF135 shows 99.0% identity over a 197aa overlap with an ORF (ORF135a) from strain A of *N.*

meningitidis:

[illegible]

30 The complete length ORF135a nucleotide sequence <SEO ID 543> is:

	1	ATGTGATACCG	CAAAAAGAAGA	CATTTTAGGA	TGGGGCTGGA	TGCTGGTGGC
	51	GGCGGCTGCG	TATTACCATTA	TGAAGCATTT	GATTAAAGAG	GATACCGCRAA
	101	AATTTGCCCT	CGGACGCGGC	GAATGTGTC	TTTGCGGCAT	GCTGTTTTCA
35	151	ACGGTTCGCG	TCCGGGCTGC	CCGGCTATTT	CCTGGGGGCA	CTCTCCGCGA
	201	CGCCCATATG	AAAAACCACT	TAAACGCGAT	TATGGTGGCG	ACGGGGCGGA
	251	TGCTGTCTGT	GTTTTACGCG	GTAAACGAG	TGCGTTTCGG	CACGGCGGTT
	301	ACCTCTGAGT	ACACCTCGTC	GATTTTITTT	GGCGATTATT	CTCTCTGAT
	351	TTTGAAGAA	CGGATTCGCG	TTTACACGCA	GCGGTGTCCT	CTCTTGTTT
40	401	TTCGCGCGCT	GGTATATGCT	CTAATCCCT	AGCTCCCGAC	CGCTCAGGAA
	451	ACGGCGGCAC	TCCGCGGCTG	GGCGGCGCG	CGGATGTTCG	GTGGGCGCTA
	501	TTTGAAGAGT	CGGGAACCTG	CTTTGGCGGG	CGAACCCGCG	TGGGCGGTGT
	551	TGTTTACCT	TTCGCTGACA	GTCTTSCGGA	TGTCACTCGT	TGGGCGGACG
	601	CTGACCGCT	GGACACCTCT	GTCTTTTCA	TGGGACAGTT	ATCTGCTGTG
	651	CTATCGGCGT	TCCGCGCTGA	TTTCTTCACT	TGCGTAGGTT	CGCTCGTCA
45	701	AGTCCGCGCA	CGAATACGCG	GTTCGCTCG	CTCTTATAT	TTCGCGTCA
	751	TTTTCCTGCT	CGGCGGAGCG	ATTCTTCTGC	GGGGGAGG	GGGCGGAGG
	801	GGAAATATCT	GCGTATGTCA	CTATCATCT	CACGGGTGAT	TTAGACAGCA
	851	TCCGCCCCAC	TGCTTCAAAA	CACGCGCTGC	AATCCCTGTT	CCGCAAAAGA
	901	TAA				

50 This encodes a protein having amino acid sequence <SEQ ID 544>:

55

1	<u>MTAKKIDIG</u>	<u>SGWMLVAAC</u>	<u>FTIMNVLKE</u>	<u>ASAKPALGSG</u>	<u>ELVFRMLFS</u>
51	<u>TVLGAAGAVL</u>	<u>RRTDFTRPW</u>	<u>KNLNLRNMG</u>	<u>TGAMLLFLYA</u>	<u>VTHLPLATGV</u>
101	<u>TLSTYSSTVL</u>	<u>AVFSLLIKE</u>	<u>LRSVYTAQL</u>	<u>LLGFAGVLL</u>	<u>LNFSRSGQGE</u>
151	<u>TALGLAGLAG</u>	<u>AMSGWAYIK</u>	<u>RELSIAGEPG</u>	<u>RWVVFYLSV</u>	<u>GVAMSSVWAT</u>
201	<u>LTGWHTLSFP</u>	<u>SALYSLCIGV</u>	<u>SALIAGLSM</u>	<u>RAYVYGDGFT</u>	<u>VASLSYTSYVT</u>
251	<u>FSALSAFFL</u>	<u>AEEFLWFOEL</u>	<u>GMCIILSGI</u>	<u>LSISIRPTAFK</u>	<u>ORLQSLFRQR</u>
301	*				

ORF135a and ORF135-1 show 99.3% identity in 300 aa overlap:

60 orf135a.pep MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMLFSTVALGAAAVL
orf135-1 MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMLFSTVALGAAAVL

5	orf135a.pep	RRDTFRTPHWKHNLNRSVMGTGAMLLFYAVTHLPLATGVTLSTYSSTIFLAVFSFLILKE
	orf135-1	RRDXFRTPHWKHNLNRSVMGTGAMLLFYAVTHLPLATGVTLSTYSSTIFLAVFSFLILKE
	orf135a.pep	RISVYTOAVLLLGAGVLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG
	orf135-1	RISVYTOAVLLLGAGVLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG
10	orf135a.pep	WRVVFYLSVTGVAMSSVWATLTGWHTLSFSPSAVYLSGIGVSALIAQLSMTRAYKVGDKFT
	orf135-1	WRVVFYLSVTGVAMSSVWATLTGWHTLSFSPSAVYLSGIGVSALIAQLSMTRAYKVGDKFT
15	orf135a.pep	VASLSYMTVVFSAASAFFLAEFLWQEIILGMCIIISGLSSIRPTAFKQRLQSLFRQR
	orf135-1	VASLSYMTVVFSAASAFFLGEELFWQEIILGMCIIISGLSSIRPTAFKQRLQSLFRQR

Homology with a predicted ORF from *N.gonorrhoeae*

ORF135 shows 97% identity over a 201aa overlap with a predicted ORF (ORF135ng) from

20	<i>N.gonorrhoeae</i> :			
	orf135.pep		GTGAMLLFYAVTXLPLATGVTLSTYSSTIF	30
	orf135ng		STVTLGAAAVLRDRTFRTPHWKHNLNRSVMGTGAMLLFYAVTHLPLTGVTLSTYSSTIF	335
25	orf135.pep		LAVFSFLILKERISVYTOAVLLLGAGVLLNPSFRSGQETAALAGLAGGAMSGWAYLK	90
	orf135ng		LAVFSFLILKERISVYTOAVLLLGAGVLLNPSFRSGQETAALAGLAGGAMSGWAYLK	395
	orf135.pep		VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFSPSAVYLSGIGVSALIAQLSM	150
30	orf135ng		VRELSLAGEPGWRVVFYLSATGVAMSSVWATLTGWHTLSFSPSAVYLSGIGVSALIAQLSM	455
	orf135.pep		TRAYKVGDKFTVASLSYMTVVFSAASAFFLGEELFWQEIILGMCIIISAVF	201
35	orf135ng		TRAYKVGDKFTVASLSYMTVVFSAASAFFLGEELFWQEIILGMCIIISAAF	506

An ORF135ng nucleotide sequence <SEQ ID 545> was predicted to encode a protein having amino acid sequence <SEQ ID 546>:

40	1 MPSEKAFRRH LRTASFQGLH LHHFHQKVGK CGIIGFGIHI FPTLLPAAQG
	51 ILDIQLGLFR IDFAALAVYR RTQVDFTHTV IDGIASDAQF SEVQILRLR
	101 NIGHFTDTHL IAQARRFIAD FGNIRFMRRG EAKTFCRCFR FDGIDGIHGD
	151 FRQCGHINRL APGKDCRNGK RDKVFFHTRH YNQVCLEKFN CSARKIKFRH
	201 QKQAKTHSTS LAARFTIRFS LSQRPFMDTA KKDILGSGWM LVAAACETVM
45	251 NVLIKEASAK FALGSGELVF WRMLFSTVTI GAAVILRDT FRTPHWKHNL
	301 NRSVMGTGAM LLLFYAVTHL PLTGTVTSY TSSIFLAVFS FLILKERISV
	351 YTOAVLLLG AGVLLNPS FRSGQEPAL AGLAGGAMSG WAYLKVRELS
	401 LAGEPGWRV VFYLSATGVAM SSVWATLTGW HTLSFSPSAV LSGIGVSALI
	451 AQLSMTRAYK VGDKFTVASL SYMTVVFSA LAAFFLGEEL FWQEIILGMCII
	501 IISAAF*

Further work revealed the following gonococcal sequence <SEQ ID 547>:

50	1 AIGGATACCG CAAAAAAGA CATTTTAGGA TCGGCGTGG TCGTGGTGG
	51 GCGGCGTGC TTCACCGTTA TGAACGATTT GATTAAAGG GCATCGGCA
	101 AATTTCGCGT CGGACGCGGC GAATTGGTCT TTGGGCGCAT GCTGTTTCA
	151 ACCGTATACCG TCGGTGCTGC CGCGGTATG CGGCGGACA CCTCCGCGA
55	201 GCCCATTTGG AAAAACCACT TAAACCGCAG TATGTCGGG ACGGGGCGA
	251 TGCTGCTGCT GTTTTACGCG GTAACGCATC TGCCTTTGAC AACCGCGGT
	301 ACCCTGAGTT ACACCTCGTC GATTTTttt GCGGATATTT CTTCTGTAT
	351 TTTGAAAGAA CGGATTTCGG TTTACACGCA GCGGTCGCT GCTCTGGTT
	401 TTGCGGCGGT GGTATTGCTG CTTAATCCCT CGTTCGCGA CGGTGAGGA
60	451 CCGCGGCGAC TCGCGGGGCT GCGGCGCGC GCGATGTCG GCTGGCGTA
	501 TTTGAAAGTG CGGAACTGT CTTTGGCGGG CGAACCGCGG TGGCGGCTG
	551 TGTTTTACCT TTCGCAACC GCGTGGCGA TGTCGTGcgt ttggcgagc
	601 Ctgaccggct ggCACAccct GTCCTTtcca tcggcagtt ATCtGtGGG

651 CATCGGCGTG tcggcgCtga TTGCCCACT GtcgatgAcg cGGCctaca
 701 agCTCGGCA CAATTTCAGG GTGCGCTGCG ttccctATk gacgtgcTC
 751 TTTTCGCGCC TGCTCTGCGC ATTTTTCCTg ggcgaagagc tttTcgCT
 801 GGAATACTC GGTATGTGCA TCATTatccT CAGCGGCATT TTGAGCAGCA
 5 851 TCGGCCCAT TGCTTCAA CAGCGCTGC AAGCCCTCTT CCGCAAGA
 901 TAA

This corresponds to the amino acid sequence <SEQ ID 548; ORF135ng-1>:

1 MDTAKKDILG SCWMLVAAC FTVMNVLKE ASAKFALGSG ELVFWRMFLS
 51 TVTLGAAAVL RRDFTFRPHW KNHLNRSVMG TGAMLLLEYA VTHLPLTTGV
 10 101 TISYTSIFL AVFSFLIKE RISVYQAVL LLGFAGVLL LNPSFRSGQE
 151 PAALAGLAGG AMSGWYLVK RELSLAGEPG WRVVFYLSAT GVAMSSVWAT
 201 LTGWHTLSFP SAVYLSGIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV
 251 FSALSAFFL GEELFWQEL GMCIIILSGI LSSIRPIAFK QRLQALFRQR
 301 *

15 ORF135ng-1 and ORF135-1 show 97.0% identity in 300 aa overlap:

orf135ng-1.pep MDTAKKDILGSCWMLVAACFTVMNVLKEASAKFALGSGELVFWRMFLSTVTLGAAAVL
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 orf135-1 MDTAKKDILGSCWMLVAACFTIMNVLKEASAKFALGSGELVFWRMFLSTVALGAAAVL
 20 orf135ng-1.pep RRDFTFRPHWKNHLNRSVMGCTGAMLLFYAVTHLPLTTGVTLSYTSISIFLAVFSFLIKE
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 orf135-1 RDXFRTPHWKNHLNRSVMGTGAMLLFYAVTHLPLATGVTLSYTSISIFLAVFSFLIKE
 25 orf135ng-1.pep RISVYQAVLLLG FAGVLLLNPSFRSGQEPALAGLAGGAMSGWAYLVKRELSLAGEPG
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 orf135-1 RISVYQAVLLLG FAGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLVKRELSLAGEPG
 30 orf135ng-1.pep WRVVFYLSATGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSMTRAYKVGDKFT
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 orf135-1 WRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSMTRAYKVGDKFT
 35 orf135ng-1.pep VASLSYMTVVSALSAFFLGEELFWQELGMCIIILSGILSSIRPIAFKQRLQALFRQR
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 orf135-1 VASLSYMTVVSALSAFFLGEELFWQELGMCIIILSGILSSIRPIAFKQRLQSLFRQR

35 Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 66

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 549>:

40 1 ATGAAGCGGC GTATAGCCGT CTTCGTCCTG TTCCGCGAGA TAATCCGAGT
 51 TTTGGGACAA CTGTGCCGGA AAATCGTCAA TACAGTTCOG GCACATCGSA
 101 TGCTCTTCCA GATTTCGGGG ATGTCTTTTT TCTTCATACA CCAGCAATAT
 151 CTGCCCGGGA TCGCGGAAAT CGATTCCCCA TCGCGCATCG TGTTCGGTGC
 201 GCTCCTCTTC CTTCACTGCG CCGCGCATGG COTGTATGTT AAAGCGCGCG
 45 251 TAGGGGATGC CgTTGCACAC GAACATCCAG TCGCTGATGT CGTCAACOGG
 301 AACGCAACGC CTTCGCGCTT GTTCGACATT GTCAGTTTG CCsGGTTCAT
 351 TGTTCAGCAC ACCGTAAATA TAAAGACCGT CAAATAAAT ATCGTGCATC
 401 CACATATGTT CGCAAAATTC CGCGTCTTCG CGCTCTTGG AAAAAGGGAC
 451 TTTGACCATG GCAAAATCCA AGCGGGAAT AATGCGCGCG CGTTCACAAA
 50 501 AAAGCTGCGG CCAAAATAT TGAATGTTT TACGGGCGCG TTGCTCGGGA
 551 OGGTTTACCG GTTCGTCTGC CTGTCTTACA TAATAATAGA GGAATCGCC
 601 CATCATATCT GCTCTCAAC GTGACGGTA TCTGTTTGA CTTACTGCG
 651 GCTTTCgtcC KTCGGCATCC GATTGGAAT TGAAGAATTC mmrwyATTGG
 701 GAATAG

55 This corresponds to the amino acid sequence <SEQ ID 550; ORF136>:

1 MKRRIAVFVL FPQIRVLGQ LLPKIVTVF AHRMLFQIFG MFFFIHQYQ
 51 LPGIAEIDSP CGIVFGALLF RHLPAHCLYG KAAVGDAVAH EHPVADVNR

```

101 NANAFALFDI GQFAXFIVQH TVNIKTVKIN IVDPHMFANF AVFAVLEKRD
151 FDHGKIQGGN NAAAFKKLA PKIFECFTGA FVGTVYRFVC LFYIINDGIA
201 HHSAFQVRVY LFAPYCGFLF SASDSDLKSS XXSE*

```

Further work revealed the complete nucleotide sequence <SEQ ID 551>:

```

5      1 ATGATGAAGC GCGGTATAGC CGTCTTCGTC CTGTTCCCGC AGATAATCCG
      51 AGTTTGGGA CCACTGTTGC CGAAAATCGT CAATACAGT CCGGCACATC
     101 GGATGCTCTT CCAGATTTC CCGGATGTTCT TTTTCTTCAT ACACAGACAA
     151 TATCTGCGCG GGATGCGCGA AATCGATTCC CCATGCGCGA TCGGTGTCGG
     201 TCGGCTCCTC TTCCGTCATC TGCCCGCGCA TTGCGTGTAT GGTAAAGCGC
     251 CCGTAGGGGA TGCCGTTGCA CACGAACATC CAGTCGCTGA TGTCTCAAC
     301 CGGAACGCCA ACGCTTTCGC CTGTGTCGAC ATTGGTCAGT TCGCGGGTTC
     351 CATTGTTTCA CACACCGTAA ATATAAAGAC CGTCAAAATA AATATCGTCG
     401 ATCCACATAT GTTCGCAAAAT TTCCGCGTCT TCGCCGCTCT GGAATAAAGG
     451 GACTTTGACC ATGGCAAAAT CCAAGGCGGA AATAATCGCG CGGCGTTCCT
     501 AAAAAAGCTC GCGCCAAAAA TATTTGAATG TTTTACGGCG CGGTTCGTCG
     551 GCACGGTTTA CCGGTTTCGC TGCCGTTTCT ACATAATAAA TGACGGAATC
     601 GCCCATCATT CTGCTCCTCA ACGTGTACGG TATCTGTTTG CACTTACTGT
     651 CGGCTTTCTG CTTTCGGCAT CCGATTTCGA TTTGAAAAGT TCCAAATATT
     701 CGGAATAG

```

20 This corresponds to the amino acid sequence <SEQ ID 552; ORF136-1>:

```

      1 MKKRRIAVFV LFPQIIRVLG QLLPKIVNTV PAHRMLFOIF GNMFFFFIHQ
     51 YLPGIAEIDS PCGIVFGALL FRHLPAHCLY GKAAVGDAVA HEHPVADVNV
    101 RNANAFALFD IGQFAGFIVQ HTVNIKTVKI NIVDPHMFAN FAVFAVLEKR
    151 DFDHGKIQGG NAAAFPKKL APKIFECFTG FVGTVYRFVC CLFYIINDGI
    201 HHSAFQVRVY LFAPYCGFLF SASDSDLKSS SKYSE*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF136 shows 71.7% identity over a 237aa overlap with an ORF (ORF136a) from strain A of *N. meningitidis*:

```

30      10      20      30      40      50      59
    orf136.pep MKKRRIAVFVLPQIIRVLGQLLPKIVNTVPAHRMLFOIFGMFFFFIHQQLPGIAEIDS
              | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
    orf136a    MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQXFGMFFFFIHQQLPGIAEIDS
              10      20      30      40      50      60
35      60      70      80      90      100     110     119
    orf136.pep PCGIVFGALLFRHLPAHCLYGKAAVGVDAVAHEHPVADVNVNRNANAFALFDIGQFAXFIVQ
              : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
    orf136a    PCGIVFGTLLFRHXSTHLYGKAAVGVNAVAHEHPVADVNVNRNANAFALFDIGQFAGFIVQ
              70      80      90      100     110     120
40      120     130     140     150     160     170     179
    orf136.pep HTVNIKTVKINIVDPHMFANFAVFAVLEKRDVDHGKIQGGNNAAPKPLAPKIFECFTG
              | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
    orf136a    HAINVKTVKINIVDPHMFANFAVFAVLEKRALTKMAKSKXXXMRRSQKSRQKVINLVLA
              130     140     150     160     170     180
45      180     190     200     210     220     230
    orf136.pep AFVGTVYRFVCLFYIINDGIAHH---SAFQVRVYLFAPYCGFLSASDSDLKSSXXSEX
              : | | : : : | | | | | | | | | | | | | | | | | | | | | | | |
    orf136a    R---SPAREFTGLSACSTXXMTESPIISAFQVRVYLFAPYCGFLSASDSDLKSSXXSEX
              190     200     210     220     230

```

The complete length ORF136a nucleotide sequence <SEQ ID 553> is:

```

55      1 ATGATGAAGC GCGGTATAGC CGTCTTCGTC CTGCTCATGC AGAAAATCCG
      51 GATTTGGGA CCACTGTTGC CGAAAATCGT CAATACAGT CCGGCACATC
     101 GGATGCTCTT CCAGATNTTC GGGATGTTCT TTTTCTTCAT ACACAGACAA
     151 TACCTGCGCG GGATGCGCGA AATCGATTCC CCATGCGCGA TCGGTGTCGG
     201 TACGCTCCTC TTCCGTCATC NGTCCACGCA TTGCTGTAT GGTAAAGCGC
     251 CCGTAGGGRA TGCCGTTGCA CACGAACATC CAGTCGCTGA TGTCTCAAC

```

5
301 CGGAACGCAA ACGCTTTCGC CTTGTCGAC ATTGGTCAGT TCGCCGGGTT
351 CATGTGTTGAC CTGCGCAATA ATATCAAGAC CGTCAAAATG
401 ATCCACATAT GTTCCCAAT TCGCCGCTCT GGAAALAAAG
451 GCTTTGACCA TGGCAAAATC TAAGGNGNNA NMGATCGCGC GCGCTTCCCA
501 AAAAAGCTCG CGCCAAATAT ATTGAATGT TTTGCGGCGC CGTTCGCGCG
551 CACGGTTTAC CGGTTTCTCT GCGTGTCTTA CATATAAAT GACGGAATCG
601 CCCATCATAT CTGCTCCTCA ACGTGTACGG TATCTGTTG CACCTTACTG
651 CGGCTTCTCG CTTTCGSCAT CCGATTCGGA TTGAAAAAGT TCCAAATATT
701 CGGAATAG

10 This encodes a protein having amino acid sequence <SEQ ID 554>:

1 MMKRRIAVFV LLMQKIRILG QLLPKIVNTV PAHRMLFQXF GMFFFFIHQQ
51 YLPGIAEIDS PCGIVFGTLL FRHXSTHCLY GKAAVGNAVA HEHPVADVNV
101 RNANAFALFD IQGFAGFIVQ HAINVKT VKI NIVDPHMFAN FAXFAVLEKR
151 ALTMAKSKXX XMRRRSQKSS RQKYLNVLRA RSPARFTGLS ACST**MTES
15 PIISAPQVRV YLFAPYCGFL PSASDSLKS SKYSE*

ORF136a and ORF136-1 show 73.1% identity in 238 aa overlap:

		10	20	30	40	50	60
orf136a.pep	MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQXFGMFFFFIHQQYLPGLIAEIDS						
orf136-1	MMKRRIAVFVLLFPQIRIVLQGLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLPGLIAEIDS						
		10	20	30	40	50	60
orf136a.pep	PCGIVFGTLLFRHXSTHCLYGKAAVGNAVAHEHPVADVNVNRNANAFALFDIQGFAGFIVQ						
orf136-1	PCGIVFGALLFRHLPAHCLYGKAAVGDAVAHEHPVADVNVNRNANAFALFDIQGFAGFIVQ						
		70	80	90	100	110	120
orf136a.pep	HAINVKT VKI NIVDPHMFAN FAXFAVLEKR ALTMAKSKXX XMRRRSQKSRQKYLNVLRA						
orf136-1	HTVNIKT VKI NIVDPHMFAN FAXFAVLEKR DFDHGKIQGGNNAAFPKKLAPKIFECFTG						
		130	140	150	160	170	180
orf136a.pep	R---SPARFTGLSACSTXXMTESPIISAPQVRVYLFAPYCGFLPSASDSLKSSKYSEX						
orf136-1	AFVGTVYRVCFLFYIINDGIAHH---SAPQVRVYLFAPYCGFLPSASDSLKSSKYSEX						
		190	200	210	220	230	
orf136a.pep							
orf136-1							

Homology with a predicted ORF from *N.gonorrhoeae*

ORF136 shows 92.3% identity over a 234aa overlap with a predicted ORF (ORF136ng) from

N.gonorrhoeae:

orf136.pep	MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLPGLIAEIDS	59
orf136ng	MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLPGLIAEIDS	60
orf136.pep	PCGIVFGALLFRHLPAHCLYGKAAVGDAVAHEHPVADVNVNRNANAFALFDIQGFAGFIVQ	119
orf136ng	PCGIVFGTLLFRHLPAHCLYGKAAVGDAVAHEHPVADVNVNRNANAFALFDIQGSAGFIVQ	120
orf136.pep	HTVNIKT VKI NIVDPHMFAN FAXFAVLEKR DFDHGKIQGGNNAAFPKKLAPKIFECFTG	179
orf136ng	HTVNIKT VKI NIVDPHMFAN FAXFAVLEKR DFDHGKIQGGNNAAFPKKLAPKIFECFTG	180
orf136.pep	AFVGTVYRVCFLFYIINDGIAHHSAPQVRVYLFAPYCGFLPSASDSLKSSXXSE	234
orf136ng	AFAGTVYRVCFLFYIINDGIAHTAPQVRVYLFAPYRGLFPLPSASDSLKSSKYSE	235

The complete length ORF136ng nucleotide sequence <SEQ ID 555> is:

1 ATGATGAAGC GCGTATAGC CGTCTCGTC CTGCTCATGC AGAAATCCG
51 GATTTTGGGA CCACTGTTGC CGAAATCGT CAATACAGTT CCGGCACATC

```

101  GGATGCTCTT  CCAATTTTC  GGGATGTCTT  TTTTCTTCAT  ACACGGCGAA
151  TACCTGCCCG  GGATCGCGCA  AATCGATTCG  CAGGCGGTA  TCCTGTTCGG
201  TACGCTCTTC  TTCGCTCATC  TGTCCGCGCA  TTGCTCTGAC  GGTAAACGCC
251  CCGTAGGGGA  TGCCGTTCGA  CACGAACATC  CAGTCTGCTA  TGTCCGCAAC
301  CGGAACGCCA  ACGCTTTCGC  CTGTTTCGAC  ATTGTTCAGT  CGCCCGGGTT
351  CATTGTTCAG  CACACCGTAA  ATATAAAGAC  CGTCAAAATA  AATATCGTCG
401  ATCCACATAT  GTTCGCAAA  TTCGCCGTCT  TCGCCGTCTT  GGAATAAAGG
451  GACTTTGACC  ATGGCAAAAT  CCAAGCGGGA  AATATGCGG  CGCGCTTCCC
501  AAAAAGCTC  GCGCCAAAG  TATTTGAATG  TTTTACGGCG  GCGTTCGCCG
551  GCACGGTTTA  CGCGTTCGTC  TGCCTGTTCT  ACATAATAAA  TGACGGAATC
601  GCCCATCATA  CTGCTCTCA  ACGTTACGG  TATCTGTTTG  CACCTTACCG
651  CGGTTTTCTA  CTTCCGGCAT  CCGATTCCGA  TTTGAAAGT  TCCAAATATT
701  CGGAATAG

```

This encodes a protein having amino acid sequence <SEQ ID 556>:

```

151  1  MMKRIIAVFV  LLMQKIRILG  QLLPKIVNTV  PAHRMLFQIF  GMFFFFIHRQ
51  YLPGIAEIDS  PGGIVFGTLL  FRHLSAHCY  GKAAYGDAVA  HEHPVADVAN
101  RNANAFALFD  IGQSAGFIVQ  HTVNIKTVKI  NIVDPHMFAN  FAVFAVLEKR
151  DFDHGKIQGG  NNAAPFKKL  AFKVFECFTG  AFAGTVYRFF  CLEYIINDGI
201  AHHTAPQVR  YLFAPYRGFL  PFASDSDLKS  SKYSE*

```

20 ORF136ng and ORF136-1 show 93.6% identity in 235 aa overlap:

```

orfl36ng      MMKRIIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFIHRQYLPGIAEIDS
                |||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:
orfl36-1      MMKRIIAVFVLLFPQIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHRQYLPGIAEIDS
25 orfl36ng      PGGIVFGTLLFRHLSAHCYLGKAAYGDAVAHEHPVADVNRNANAFALFDIGQSAGFIVQ
                |||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:
orfl36-1      PGGIVFGALLFRHLSAHCYLGKAAYGDAVAHEHPVADVNRNANAFALFDIGQSAGFIVQ
30 orfl36ng      HTVNIKTVKINIVDPHMFANFAVFAVLEKRFDFHGKIQGGNNAAPFKKLAPKVFECFTG
                |||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:
orfl36-1      HTVNIKTVKINIVDPHMFANFAVFAVLEKRFDFHGKIQGGNNAAPFKKLAPKVFECFTG
orfl36ng      AFAGTVYRFFVCLFYIINDGIAHHTAPQVRVYLFAPYRGFLPFASDSDLKSSKYSEX
                ||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:
35 orfl36-1      AFVGTVYRFFVCLFYIINDGIAHHTAPQVRVYLFAPYCGFLPSASDSDLKSSKYSEX

```

Based on the presence of the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 67

40 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 557>:

```

1  ATGGAAATA  TGGTAACGTT  TTCAAAAATC  AGACCGCTTT  TGCCAATCGC
51  CGCGCGCCG  TTGCTTCCG  CC.TGCGGAC  GCGCGGAAT  AATGCTGTCC
101  CAAAGCCGG  GCAACCCGCC  AAACCCGCCG  CAGTGGTCGG  TTTGCGCACT
151  GGTGGCGGG  CATCTAAAG  ATTGCGCAT  GTAGGTATTA  TTAAGTTTTT
45 201  GAAAGAAAC  GGTATTCCTG  TGAAGTGGT  TACCGGCACC  TCGCGAGTTT
251  CGATTGTGG  CAACCTTTT  GCAATCGGTA  TGTCCGCCGA  CGCGCTCGAA
301  TTGGAAGCC  AAATTTTAG  CAAACCGAT  TTGCTGATT  TAACCTTTG
351  CACCAATGG  TTTATCAAG  GCGCAAGCT  GCAAAATTA  ATCAACGAA
401  AACTCCGCG  CATGCAATT  CAGCAGTTT  CCATCAAA  TGCGGCC..

```

50 This corresponds to the amino acid sequence <SEQ ID 558; ORF137>:

```

1  MENMVTFSKI  RPLLAIAAAA  LLAAXRTAGN  NAVRKPVQTA  KPAVVGLAL
51  GGGASKGF  VGLIKVLKEN  GIPVKVVTGT  SAGSIVGNLF  ASGMSFDRL
101  LEAEILGKTD  LVDLTLSTNG  FIKGAKLQNY  INRKLGRMQI  QQFFIKFAA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 559>:

```

1  ATGGAATAA TGTTAAGCTT TCAAAAATC AGACCGCTTT TGGCAATCGC
51  CGCGCGCGCG TGCTTGGCG GCGCGGACAT AATGCTGCGC
101  GCAGCGCGCT GCAAAACCGC AAACCGCGCG CAGTGTGCGG TTTGGCCTC
151  GGTGGCGCGC CATCTAAAGG ATTGGCCAT GTAGGTATTA TTAAGGTTTT
5  201  GAAAGAAAC GGTATTCTGT TGAAGTGGT TACCGGCACA TCGGCAGGTT
251  CGATTGTGCG CAGCCTTTTT GCATCGGGTA TGTGCGCGA CGCCTCGAA
301  TTGGAAGCGG AAATTTTAGG CAAACCGAT TTTGTCGATT TAACCTTGTC
351  CACCACTGGT TTTATCAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA
10  401  AAGTCGCGCG CAGGCAGATT CAGCAGTTTC CCATCAAAAT TCGCCGCGTT
451  GCTACTGATT TTGAAACCGG CAAGGCGGTC GCTTTCAATC AGGGGAATGC
501  CGGCGAGGCT GTGCGCGCTT CGCGCGCAT TCCTCAATGT TTCCAAACCG
551  TTATCATCGG CAGGCATACA TATGTTGAG CGGCTCTGTC GCAGCCCGTG
601  CCGCTCAGTG CGCGCGCGCG CAGGCGCGCT AATTTCGTGA TTGCGCTCGA
651  TATTTCGCGC CGTCCGGGCA AAAACATCAG CCAAGGTTTC TTCTCTTATC
15  701  TCGATCAGAG GCTGAACGTA ATGAGCGGTT CTGCGTTGCA AAATGAGTTG
751  GGGCAGGCGG ATGTGGTTAT CAAACCGCAG GTTTTGGATT TGGGTGCGAT
801  CGGCGGATTC GATCAGAAAA AAGCGCGCAT CCGGTTGGGT GAGGAGGCTAT
851  CAGCTGCGCG ATTGCTGAAA ATCAACCGCA AACTGCGCGC
901  TGA

```

20 This corresponds to the amino acid sequence <SEQ ID 560; ORF137-1>:

```

1  MENMVTFSKI RPLLAIAAAA LLAACGTAGN NAVRKPVQTA KPAAVVGLAL
51  GGGASKGFAH VGIKVLKEN GIPVKVVTGT SAGSIVGSLF ASGMSPDRL
101  LEAEILGKTD LVDLTLSTG FIKGKQLQNY INRKVGGRQI QQPIKEAAV
25  151  ATDFETGKAV AFNQGNAGQA VRASAAIPNV FQPIVIGRHT YVDGGLSPV
201  PVSAARRQGA NFVIAVDISA RPKNII SQGF FSYLDQTLNV MSVSALQNEL
251  GQADVVIKQP VLDLGA VGGF DQKRAIRLG EEAARAALPE IKRKLAAARY
301  *

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

30 ORF137 shows 93.3% identity over a 149aa overlap with an ORF (ORF137a) from strain A of *N. meningitidis*:

```

      10      20      30      40      50      60
orf137.pep  MENMVTFSKIRPLLAIAAAAALAAAXRTAGNNAVRKPVQTA KPAAVVGLALGGGASKGFAH
35  orf137a  MENMVTFSKIRPLLAIAAAAALAAAGCTAGNNAARKEPVQTA KPAAVVGLALGGGASKGFAH
      10      20      30      40      50      60
      70      80      90      100     110     120
40  orf137.pep  VGIKVLKENGIPVKVVTGT SAGSIVGNLFASGMSPDRLLEAEILGKTDLVDLTLSTNG
orf137a  VGIKVLKENGIPVKVVTGT SAGSIVGSLFASGMSPDRLLEAEILGKTDLVDLTLSTNG
      70      80      90      100     110     120
45  orf137.pep  FIKGAKLQNY INRKLRGMIQQPPIKFAA
orf137a  FIKGKQLQNY INRKVGGRRIQQPIKFAA
      130     140     149
      130     140     150     160     170     180

```

The complete length ORF137a nucleotide sequence <SEQ ID 561> is:

```

50  1  ATGGAATAA TGTTAAGCTT TCAAAAATC AGACCGGCTTT TGGCAATCGC
51  CGCGCGCGCG TGCTTGGCG GCGCGGACAT AATGCTGCGC
101  GCAGCGCGCT GCAAAACCGC AAACCGCGCG CAGTGTGCGG TTTGGCCTC
151  GGTGGCGCGC CATCTAAAGG ATTGGCCAT GTAGGTATTA TTAAGGTTTT
55  201  GAAAGAAAC GGTATTCTGT TGAAGTGGT TACCGGCACA TCGGCAGGTT
251  CGATAGTCGG CAGCCTTTTT GCATCGGGTA TGTGCGCGA CGCCTCGAA
301  TTGGAAGCGG AAATTTTAGG TAAACCGAT TTTGTCGATT TAACCTTGTC
351  CACCACTGGT TTTATCAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA
401  AAGTCGCGCG CAGGCAGATT CAGCAGTTTC CCATCAAAAT TCGCCGCGTT
451  GCTACTGATT TTGAAACCGG CAAGGCGGTC GCTTTCAATC AAGGGGAATGC
60  501  CGGCGAGGCT GTGCGCGCTT CGCGCGCAT TCCTCAATGT TTCCAAACCG
551  TTATCATCGG CAGGCATACA TATGTTGAG CGGCTCTGTC GCAGCCCGTG

```

5
 601 CCGGTCACTG CCGCCCGGGC GCANGNNNG NATNTCGTA TTGCCGTCGA
 651 TATTTCGGC CFTCCGAGCA AATCATCAG CCAAGCTTC TTCTCTTATC
 701 TCAGTCAGAC GCTGAACGTA ATGAGCGTTT CCGCGTTCGA AATGAGTTG
 751 GGGCAGCGGG ATGTGGTTAT CAACCCGAG GTTTGGATT TGGGTGCACT
 801 CCGCGATTTC GATCAGAAAA AACCGCCAT CCGTTGGGT GAGGAGCGAG
 851 CACGTGCGGC ATTGCTGAA ATCAACGCA AACTGCGGC ATACCGTTAT
 901 TGA

This encodes a protein having amino acid sequence <SEQ ID 562>:

10
 1 MENMVTFSKI RPLLAIAAAA LLAACGTAGN NAARKPVQTA KPAAVVGLAL
 51 GGGASKGFH VGIKVLKEN GIPVKVVTG SAGSIVGSLF ASGMSPDRL
 101 LEAEILGKTD LVDLTSTSG FIKGEKLQNY INRKVGRRRI QQFPIKFAAV
 151 ATDFETGKAV AFNQGNAGQA VRASAAIPNV FQPVIIGRHT YVDGGLSQPV
 201 PVSAARXXX XXVIAVDISA RPSKNISQGF FSYLDQTLNV MSVSALQNEL
 251 GQADVVIKQV VLDLGA VGGF DQKKRAIRLG EEAARAALPE IKRKLAAARY
 15 301 *

ORF137a and ORF137-1 show 97.3% identity in 300 aa overlap:

20
 orf137a.pep MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAARKPVQTA KPAAVVGLALGGGASKGFH
 orf137-1 MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAARKPVQTA KPAAVVGLALGGGASKGFH
 25
 orf137a.pep VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLLEAEILGKTDLVDLTSTSG
 orf137-1 VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLLEAEILGKTDLVDLTSTSG
 30
 orf137a.pep FIKGEKLQNYINRKVGRRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV
 orf137-1 FIKGEKLQNYINRKVGRRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV
 35
 orf137a.pep FQPVIIGRHTYVDGGLSQPVVSAARXXXVIAVDISARPSKNISQGFPSYLDQTLNV
 orf137-1 FQPVIIGRHTYVDGGLSQPVVSAARQQANFVIAVDISARPSKNISQGFPSYLDQTLNV
 orf137a.pep MSVSALQNELGQADVVIKQVLDLGA VGGF DQKKRAIRLGEEAARAALPEIKRKLAAARY
 35 orf137-1 MSVSALQNELGQADVVIKQVLDLGA VGGF DQKKRAIRLGEEAARAALPEIKRKLAAARY

Homology with a predicted ORF from *N.gonorrhoeae*

ORF137 shows 89.9% identity over a 149aa overlap with a predicted ORF (ORF137ng) from *N.gonorrhoeae*:

40
 orf137.pep MENMVTFSKIRPLLAIAAAAALLAAXRTAGNNAARKPVQTA KPAAVVGLALGGGASKGFH 60
 orf137ng MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAARKPVQTA KPAAVVGLALGGGASKGFH 60
 45
 orf137.pep VGIKVLKENGIPVKVVTGTSAGSIVGNLFASGMSPDRLLEAEILGKTDLVDLTSTSG 120
 orf137ng VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLLEAEILGKTDLVDLTSTSG 120
 orf137.pep FIKGEKLQNYINRKVGRRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV 149
 50 orf137ng FIKGEKLQNYINRKVGRRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV 180

The complete length ORF137ng nucleotide sequence <SEQ ID 563> is:

55
 1 ATGGAATAA TGTTAACGTT TTCAAAAATC AGATCATTTT TGGAATCGC
 51 CGCGCCGCGC TTGCTTGGCG CTGCGGTAC GCGGGGAAC AATGCGCGCC
 101 GCAAGCCGGT GCAAAACCGC AAACCCGCCG CAGTGGTCGC TTTGGCAGCTC
 151 GGTGGCGGCG CATCTAAAGG ATTTGGCCAT ATAGGAATGT TTAAGGTTT
 201 GAAAGAAAC GGTATTCCTG TGAAGTGGT TACCGGCACA TCGCGAGTT
 251 CGATAGTCGG CAGCCTTTT GCATCGGGTA TGTGCGCCGA CGCCTCGAA
 301 TTGAAGCCG AGATTTTAGG TAAACCGAT TTAGTCGATT TAACTTGTG
 351 CACCAAGTGT TTTATCAAG GCGAAAGCT GCAAAATAC ATCAACCGAA
 60 401 AAGTCGGCG CAGGCAGATT CAGCAGTTT CCAATCAATT TGCCGCGGTT
 451 GCCACTGATT TTGAACCGG CAAGGCCGTC GTTTCAATC AAGGGAATGC

501 CGGGCAGGGG GTTCGTGCTT CCGCCGCAT TOCCAATGTG TTCCAGCCAG
 551 TCATCTCGG CAGGCACAAA TATGTGACG CGCGTCTGTG GCAGCCCGTG
 601 CCGCTCAGTG CCGCTCGCGG CGAGCGGGCG AATTCGCTGA TTCCGCGTGA
 651 TATTTCCGCA CGTCCAGCA AAAATGTCGG TCAAGTTTTC TTCTCTATTC
 701 TCGATCAGAC GCTGACGCTG ATGAGCGTTT CCGTGTGCA AAACAGTGTG
 751 gggcAGCGGG ATGTGGTTAT CAAACCGCag gTTTGGATT TGGGTGCAAT
 801 CGCGGAGTTC GATCAGAAAA AGCGGCCCAT CCGGTTGGG GAGGAGGCAG
 851 CACGTGCCGC ATTGCCTGAA ATCAACGCA AACTGCGCGC ATACCGTTAT
 901 TGA

10 This encodes a protein having amino acid sequence <SEQ ID 564>:

1 MENMVTFSKI RSFLAIAAAA LLAACGTAGN NAARKPVQTA KPAAVVALAL
 51 GGGASKGFAH IGIVKVLKEN GIPVKVVTGT SAGSIVGSL ASGMSPDRL
 101 LEAEILGKTD LVDLTLSTSG FIKGEKLQNY INRKVGGRQI QQFFIKFAAV
 151 ATDFETGKAV AFNQGNAGQA VRASAAIPNV FQPVIIGRHK YVDGGLSQV
 201 PVSAARRQGA NFVIAVDISA RPSKNVQGF FSYLDQTLNV MSVSLQNEL
 251 GQADVVIKPV VLDLGAVGGF DQKKRAIRLG EEAARAALPE IKRKLAAARY
 301 *

ORF137ng and ORF137-1 show 96.0% identity in 300 aa overlap:

20 orf137ng MENMVTFSKIRSFLAIAAAA LLAACGTAGNNAARKPVQTA KPAAVVALALGGGASKGFAH
 orf137-1 MENMVTFSKIRFLAIAAAA LLAACGTAGNNAARKPVQTA KPAAVVALALGGGASKGFAH
 orf137ng IGIVKVLKENGIPVKVVTGTSAGSIVGSLASGMSPDRLLEAEILGKTDLVDLTLSTSG
 25 orf137-1 VGIKVLKENGIPVKVVTGTSAGSIVGSLASGMSPDRLLEAEILGKTDLVDLTLSTSG
 orf137ng FIKGEKLQNYINRKVGGRQIQFFIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV
 orf137-1 FIKGEKLQNYINRKVGGRQIQFFIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV
 30 orf137ng FQPVIIGRHKYVDGGLSQVPVSAARRQGANFVIAVDISARPSKNVQGGFFSYLDQTLNV
 orf137-1 FQPVIIGRHTYVDGGLSQVPVSAARRQGANFVIAVDISARPGKNISQGGFFSYLDQTLNV
 35 orf137ng MSVSLQNELGQADVVIKPVQVLDLGAVGGFDQKKRAIRLGEAARAALPEIKRKLAAARY
 orf137 MSVSLQNELGQADVVIKPVQVLDLGAVGGFDQKKRAIRLGEAARAALPEIKRKLAAARY

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 68

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 565>:

45 1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CTTTGGGAA CGCCATGCA
 51 CATCTCGTGTG ACAGCCCTCG TCAATGCTCT CCGCTGCTG CCGCTTCTC
 101 GTCTGCACAC GCTGGGAAC CGGCTCGGAC ATCTCGGCTT TACCTTTTA
 151 AAGGAAGACC GCGCGGCGAT CGTCGCCmAT ATCGGCGAGG CGGCTTTGA
 201 CCCGAGCCCC AAAACGGTCA AAGCGGTTTT TCGCGAAGCG GCAAAAGGCG
 251 GTTTGGAACT TCGCCCGCGG TTTTTCAGAA AACCGGAAGA CATAGAAAC
 50 301 ATGTCARAAG CGGTACACGG CTGGGAACAT GTGACGAGG CTTTGGACAA
 351 ACACGAAGGG CTGCTATTC. .

This corresponds to the amino acid sequence <SEQ ID 566; ORF138>:

1 MFRQLRFLFP PLRTAMHILL TALLKCLSL PLSCHLTLGN RLGLHAFYLL
 51 KEDRARI VAX MRQAGNLPD RTVKAVFAET AKGGLLELAPA FFRKPEIDET
 55 101 MFKAVHGEH VQALOKHEG LLF

Further work revealed the complete nucleotide sequence <SEQ ID 567>:

```

1   ATGTTTCGTT  TACAATTCAG  GCTGTTCCG  CCTTTCGAA  CCGCATGCA
51  CATCCTGTGT  ACCGCCCTGC  TCAAATGCT  CTCCTGCTG  CGGCTTTCCT
101 GTCTGCACAC  GCTGGGAAC  CGGCTCGAC  ATCTGGCGTT  TTACCTTTTA
151 AAGGAAGACC  GCGCGCGCAT  CGTGCCTAT  ATGCGGCAG  CGGCTTTGAA
201 CCCCAGCCCC  AAAACGGTCA  AAGCGGTTT  TCGCGAAGC  GCAAAGCGG
251 GTTTGGAAC  TGCCCCCGCG  TTTTTCAGAA  AACCGAAGA  CATAGAAACA
301 ATGTTCAAAG  CGGTACACGG  CTGGGAAC  GTGACGACG  CTTTGGACAA
351 ACACGAAGG  CTGCTATTC  TCAGCCGCA  CATCGCAGC  TAGGATTGG
401 GCGCAGCTA  CATCAGCCAG  CAGCTTCCT  TCCTCGCTAC  CGCATGTAC
451 AATCGCCGA  AATCAAAGC  GATAGACAA  ATCATCGAG  CGGCGAGGT
501 TCGCGCAAA  GGAACAAAC  CGCTACACG  CATACAAGG  GTCACAAACA
551 TCATCAAAGC  CCGGTTCG  GCGAAGCAA  CCATCGTCT  GCGCGACCAC
601 GTCCCTCC  CTCGAAGG  GCGGAAGC  GTATGGGTG  ATTTCTTCG
651 CAAACCTGC  TATACATGA  CGTGGCGG  AATTTGCA  CACGTCAAAG
701 GCGTGAAG  CCTGTTTT  TGCTGGAAC  GCCTGCCTG  CGGACAAGT
751 TTCGATTGC  ACATCCGCC  CGTCCAAGG  GAATTGAAC  GCGACAAGG
801 CCATGATGC  GCGCTGTTC  ACCGCAATG  CGAATTATG  ATACGCGGT
851 TTCGAGCG  GTATCTGTT  ATGTACAA  GCTACAAAT  CGCGTAA

```

20 This corresponds to the amino acid sequence <SEQ ID 568; ORF138-1>:

```

1   MFLRQLRFLP  PLRTAMHILL  TALLKCLSL  PLSCSLHTLN  RLGHIAFYLL
51  KEDRARIVAN  MRQAGLNDF  KTVKAVFAE  AKGGLELAPA  FFRKPEDIE
101 MFKAVHGEH  VQALDKHEG  LLFITPHIG  YDLGGYISQ  QLFFLTAMY
151 KPKIKAIK  IMQAGRVGK  GKTAPTSIG  VKQIKALKS  GEATVLPLFH
201 VPSPEGEG  VWDFGKFA  YMTLAAKLA  HVKGKTLFF  CCELRPGQG
251 FDLHIFVQG  ELNGDKAHD  AVFNRAEY  IRRFTQQLF  MYNRKMP*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF138 shows 99.2% identity over a 123aa overlap with an ORF (ORF138a) from strain A of *N.*

30 *meningitidis*:

```

              10      20      30      40      50      60
orf138.pep  MFLRQLRFLPPLRTAMHILLTALLKCLSLPLSCSLHTLGNRLGHIAFYLLKEDRARIVAX
              |||
orf138a     MFLRQLRFLPPLRTAMHILLTALLKCLSLPLSCSLHTLGNRLGHIAFYLLKEDRARIVAN
              10      20      30      40      50      60
              70      80      90      100     110     120
orf138.pep  MRQAGLNDFPKTVKAVFAETAKGGLELAPFAFRKPEDIEITMFKAVHGEHVQQALDKHEG
              |||
orf138a     MRQAGLNDFPKTVKAVFAETAKGGLELAPFAFRKPEDIEITMFKAVHGEHVQQALDKHEG
              70      80      90      100     110     120
              130     140     150     160     170     180
orf138.pep  LLF
              |||
orf138a     LLFITPHIGSYDLGGYISQQLFPFLTAMYKPKIKAIKIMQAGRVGKGKTAPTSIGQ

```

The complete length ORF138a nucleotide sequence <SEQ ID 569> is:

```

1   ATGTTTCGTT  TACAATTCAG  GCTGTTCCG  CCTTTCGAA  CCGCATGCA
51  CATCCTGTGT  ACCGCCCTGC  TCAAATGCT  CTCCTGCTG  CGGCTTTCCT
101 GTCTGCACAC  GCTGGGAAC  CGGCTCGAC  ATCTGGCGTT  TTACCTTTTA
151 AAGGAAGACC  GCGCGCGCAT  CGTGCCTAT  ATGCGGCAG  CAGGCATGAA
201 TCCGAGCCCC  AAAACGGTCA  AAGCGGTTT  TCGCGAAGC  GCAAAGCGG
251 GTTTGGAAC  TGCCCCCGCG  TTTTTCAGAA  AACCGAAGA  CATAGAAACA
301 ATGTTCAAAG  CGGTACACGG  CTGGGAAC  GTGACGACG  CTTTGGACAA
351 ACACGAAGG  CTGCTATTC  TCAGCCGCA  CATCGCAGC  TAGGATTGG
401 GCGCAGCTA  CATCAGCCAG  CAGCTTCCT  TCCTCGCTAC  CGCATGTAC
451 AATCGCCGA  AATCAAAGC  GATAGACAA  ATCATCGAG  CGGCGAGGT
501 TCGCGCAAA  GGAACAAAC  CGCTACACG  CATACAAGG  GTCACAAACA
551 TCATCAAAGC  CCGGTTCG  GCGAAGCAA  CCATCGTCT  GCGCGACCAC

```

5 601 GTCCCTCC CTCAAGAAGG CGGGGAAGGC GTATGGGTGG ATTTCTTCGG
 651 CAACCTTCGC TATACCATGA CAGTTCGCGC AATATGCGCA CAGCTCAAGG
 701 GCGTCAAAAC CTTGTTTTTC TCGTGGGAC GCGTGGCTGG CGGACAAGGT
 751 TTGGATTGCG ACATCCGCCG CGTCCAAGGG GAATTGAACG CGACAAGAC
 801 CCATGATGCC GCGCTGTTC ACGCAATGC CGAATATTGG ATACGCCGTT
 851 TTCCGACGCA GTATCTGTTT ATGTACAACC GCTACAAAT CCGGTTAA

This encodes a protein having amino acid sequence <SEQ ID 570>:

1 MFRLQRLFP PLRTAMHILL TALLKCLSL LPLSCLHTLGN RLGLHAFYLL
 51 KEDRARIVAN MRQAGLNPDF KTVKAVFAET AKGGLLELAPA FFRKPEDLET
 101 MFKAVHWEH VQALDKHEG LLFITPHIGS YDLGGYISQ QLPFLTAMY
 151 KPPKIKAIKD IMQAGVRGK GKTAFTSIQG VKQIIKALRS GEATIVLPDH
 201 VPSQEGEGEG VWDVDFGKFA YMTLAAKLA HVKGKTLFF CERLPGGGQ
 251 FDLHIRPVQG ELNGDKAHDA AVFNRAEYW IRRFPTQYLF MYNRYKMP*

ORF138a and ORF138-1 show 99.7% identity over a 298aa overlap:

15 orf138a.pep MFRLQRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGLHAFYLLKEDRARIVAN
 orf138-1 MFRLQRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGLHAFYLLKEDRARIVAN
 20 orf138a.pep MRQAGMNPDPKTVKAVFAETAKGGLLELAPAFFRKPEDLETMFKAVHWEHVQALDKHEG
 orf138-1 MRQAGLNPDFKTVKAVFAETAKGGLLELAPAFFRKPEDLETMFKAVHWEHVQALDKHEG
 25 orf138a.pep LLFITPHIGSYDLGGYISQQLPFPLTAMYPKPKIKAIKIMQAGVRGKGTAPTSTIQQ
 orf138-1 LLFITPHIGSYDLGGYISQQLPFPLTAMYPKPKIKAIKIMQAGVRGKGTAPTSTIQQ
 30 orf138a.pep VKQIIKALRSGEATIVLPDHPVPSQEGEGGVWDVDFGKPAYTMTLAALAHVKGKTLFF
 orf138-1 VKQIIKALRSGEATIVLPDHPVPSQEGEGGVWDVDFGKPAYTMTLAALAHVKGKTLFF
 35 orf138a.pep CCERLPGGGQFDLHIRPVQGEELNGDKAHDAAVFNRAEYWIRRFPTQYLFMYNRYKMP
 orf138-1 CCERLPGGGQFDLHIRPVQGEELNGDKAHDAAVFNRAEYWIRRFPTQYLFMYNRYKMP

35 Homology with a predicted ORF from *N.gonorrhoeae*

ORF138 shows 94.3% identity over a 123aa overlap with a predicted ORF (ORF138ng) from

N.gonorrhoeae:

orf138.pep MFRLQRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGLHAFYLLKEDRARIVAX 60
 40 orf138ng MFRLQRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGLHAFYLLKEDRARIVAN 60
 orf138.pep MRQAGLNPDFKTVKAVFAETAKGGLLELAPAFFRKPEDLETMFKAVHWEHVQALDKHEG 120
 orf138ng MRQAGLNPDFKTVKAVFAETAKGGLLELAPAFFRKPEDLETMFKAVHWEHVQALDKHEG 120
 45 orf138.pep LLF 123
 orf138ng LLFITPHIGSYDLGGYISQQLPFPLTAMYPKPKIKAIKIMQAGVRGKGTAPTSTIQQ 180

The complete length ORF138ng nucleotide sequence <SEQ ID 571> is:

50 1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CTTTGGCAA CCGCATGCA
 5 CATCCTGTTG ACCGCCCTGC TCAAAATGCTT CTTCCGTGCT TCGCTTCTCT
 101 GTCTGCACAC GCTGGGAAC CCGCTGGGAC ATCTGGCGTT TTACTTTTAA
 151 AAGGAAGACC GCGCGCGCAT CGTCCCAAT ATGCGGACG CGGCTTGAA
 201 CCGCGACACG CAGACCGTCA AAGCCGTTT TCGCGAACG GCAAAATGCG
 55 251 GTTTGGAACT TGCCTCCGCG TTTTTCAAA AACCGAAGA CATCGAARCA
 301 ATGTTCAAAG CGGTACACGG CTGGGAACAC GTGCAGCAGG CTTTGGACAA
 351 GGGCGAAGGG CTGCTGTTC TACGCGCGCA CATCGCAGC TACGATTGCG
 401 GCGGACGCTA CATCAGCCAG CAGCTTCCGT TCCACTGAC CGCCATGTAC
 451 AAGCCGCGCA AATCAAAAGC GATAGACAAA ATCATCGAGG CGGCGAGGTT
 60 501 GCGCGGCAAA GGCAAAACCG cggccaccgg catACAGGG GTCAACAAAA
 551 tcatcaAGGC CTGCGCGCGG GCGGAGGCAA CCAATCATCT GCCCGACCA

601 GTCCCTTCTC CGCAGGAagg cggCGGCGTG TGGCGGGATT TTTTGGGCAA
 651 ACCTGGCTAC accATGACAC TGCGGCAAC ATTGCGCAC GTCAAGGGC
 701 TGAAGACCT GTTTTCTGCG TCGGACCGC TCGCGACGG ACAAGGCTTC
 751 GTGTTCGACA TCGCGCCGCT CCAAGGGCAA TGAACCGCA ACAAGCCCA
 801 CGATGCGGCC GTGTTCACCC GCAATACCGA ATATTGGATA CGCGCTTTTC
 851 CGACGCAGTA TGTGTTTATG TACAACCGCT ATAAAGCGC GTAA

This encodes a protein having amino acid sequence <SEQ ID 572>:

1 MFRLQFRLFP PLRTAMHILL TALLKCLSL SLSCLHTLGN RLGHLAFYLL
 51 KEDRARIVAN MRQAGLNPD TQVKAFAET AKCGLELAPA FFKPEDIET
 101 MFKAVHGW EHQALDKGEG LLFITPHIGS YDLGGYISO OLPHLTAMY
 151 KPPKIKAIK IMQAGRVRGK GKTAPTGIQ VQKIIKALRA GEATIIILPH
 201 VPSPOEGGV WADFFGKPAY TMTLAAKLAH VKGVKTLFFC CERLPDGGGF
 251 VLHIRPVOGE LNGNKAHDAA VFNRRTEYWI RRPFTQYLFM YNRYKTP*

ORF138ng and ORF138-1 show 94.3% identity over 299aa overlap:

15 orf138-1.pep MFRLQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLAFYLLKEDRARIVAN
 orf138ng MFRLQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLAFYLLKEDRARIVAN
 20 orf138-1.pep MRQAGLNPDPTQVKAFAETAKGGLLELAPAFFRKPEDIETMFKAVHGW EHVQALDKHEG
 orf138ng MRQAGLNPDPTQVKAFAETAKGGLLELAPAFFRKPEDIETMFKAVHGW EHVQALDKHEG
 orf138-1.pep LLFITPHIGSYDLGGYISQOLPFLTAMYPKPKIKAIKIMQAGRVRGKGKTAPTSGI
 orf138ng LLFITPHIGSYDLGGYISQOLPFLTAMYPKPKIKAIKIMQAGRVRGKGKTAPTSGI
 25 orf138-1.pep VQKIIKALRSGEATIVLPDHVPSPOEGGEGVWVDFGKPAYTMTLAAKLAHVGVKTLFF
 orf138ng VQKIIKALRAGEATIIILPDHVPSPGEGG--GVWVDFGKPAYTMTLAAKLAHVGVKTLFF
 30 orf138-1.pep CCERLPGGGQFDLHIRPVOGELNGDKAHDAAVFNRRAEYWI RRPFTQYLFMYNRYKMP
 orf138ng CCERLPDGGQFVLHIRPVOGELNGNKAHDAAVFNRRTEYWI RRPFTQYLFMYNRYKTP

In addition, ORF138ng is homologous to htrB protein from *Pseudomonas fluorescens*:

35 gnl|PID|e334283 (Y14568) htrB [Pseudomonas fluorescens] Length = 253
 Score = 80.8 bits (196), Expect = 9e-15
 Identities = 49/151 (32%), Positives = 79/151 (51%), Gaps = 6/151 (3%)
 40 Query: 101 MFKAVHGW EHVQALDKGEGLLFITPHIGSYD-LGGYISQOLPFLTAMYPKPKIKAIK 159
 + + V G E + + + L G + + + I T H + G + + L Y S Q P Y + PPK + KA + D
 Sbjct: 94 LVREVEGLEVLKEALASGKGVGITSHLGNWEVIMHIFYCSQCKPI---IFYRPPKKAVID 150
 Query: 160 KIMQAGRVRGKGKTAPTGIQGVQKIIKALRAGEATIIILPDHVPSPGEGGVWVDFGKFA 219
 + + + + RV + K A + + G + + I K + R G I D P E E G + + F F A
 45 Sbjct: 151 ELLRQKRVQLGNKVAASTKEGILSVIKVRKGGQGVGIPAD--PEPAESAGIFVFFATQA 208
 Query: 220 YTMTLAAKLAHVGVKTLFFCCERLPDGGGF 250
 T + -F RLPDG G +
 Sbjct: 209 LTSKFPVNLAGGKAVGVFLHALRLPDGSGY 239

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF138-1 (57kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above.

The products of protein expression and purification were analyzed by SDS-PAGE. Figure 14A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis

(Figure 14B). These experiments confirm that ORF138-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 69

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 573>:

```

5      1  ..GCGTGGTGGG CCGGCGAATC GTGGCGTGTG TTAATGGAAA GTGAAACGTG
      51  GCATGCGGTG TGGAACTACT TGCGCTTCTC GCGGCGCGCG GTGTATGCGG
     101  CAGCGGTTTT GGGTGTGGTG TATCGGCGCC CCGGCGCGCG GTCGCGTGGG
     151  ATGCGCGGCG TGATGTTTTA GCGCTTATG GTGTCGCGCG TTGTGTTTTT
     201  GCGCGGCGTG CTGCTGTCTT ATCGCAGTGC GACGGCTTCG TTGCGGTTCG
     251  TGCTGGCGAT GTATGCGCTG CTGGCGTATC CTTTGTGGC AAAAGATGTT
     301  TTATCAGCCT GGGATGCACAT GCCCGCGGAT TACGGCAGGG CCGCGCGCGG
     351  TTTGGGTGCA AACGGCTTTC AGACGGCATG CCGCATCACG TTCCCCCTCT
     401  TGAACCGCG GTTGCGGCGC GGTCTGACTT TGGCGGCGCG AACCTGCGTG
     451  GCGCAATTTG CCGCGACATT GTTCTGTGCG GCTCCGGAAT GCGACAGCCT
     501  GACGACTTTG ATTTATGCTC ATTTGGGACG CCGGGGTGAG GATATATTACG
     551  CCGCGGCGAT GGTGCTG..
  
```

This corresponds to the amino acid sequence <SEQ ID 574; ORF139>:

```

      1  ..AWSAGESWRV LMESETWHAV WNTLRFSAAA VYAAAVLGUV YAAFARRSAA
     51  MRGLMFXPFM VSPVVCVSAGV LLLYPQWTAS LPELLAMAYL LAYFFVAKDV
    101  LSAWDALPPD YGRAAAGLGA NGFQTACRIT FPLKPAIRLR GLTLAAATCV
    151  GEFATLFLS RPEWQLTTL IYAYLGRAGE DNYARAMVL..
  
```

Further work revealed the complete nucleotide sequence <SEQ ID 575>:

```

      1  ATGGATGGAC GCGTGTGGGT GGTATGGGGT GCTTTTGCCC TGCTGCGCTC
     51  GCGCTTTTTG CCGGTAATGG TCGTTCGCC TTGTGGGCGG GTGGCGCGGT
    101  ATGACGGTTT GCGGTGGCGC CCGGTGCTGT CCGATGCTTA TAGCTCAAA
    151  CTTTTGCGGT GACGCGTATT TCAGGCGAGC GCAACCTGTG TCGTGGTGCT
    201  GCTTTTGGCG GTGCTGTGCG CGTGCTGCTT GCGCGCGCTG GCGTTTCGCG
    251  GCGCGGCTTT GGTGCTGCGC CTGCTGATGC TGCTTTTGTG GATGCCAAG
    301  TTGTGCGCGG CCGTGGCGGT GCTGGCGCTG TTGCGGCGCG ACGGCGCTTT
    351  GTGGCGCGCG AGGCAGGATA CCGCGTATCT GTTGTGTGAC GGCAATGTGT
    401  TTTTCAACCT TCCTGTGTTG GTACAGGCGG CGTATCAGGG GTTTGTGCAA
    451  GTGCTTGGCG CACGGCTTCA GACGGCAGCG ACGTATGGCG CCGGGGCGTG
    501  GCGGCGGTTT TGGGACATTG AAATGCCCGT TTGCGCGCGG TGCGTTGCGG
    551  GCGGCGGTGT CTTGTCTTTT CTGTATTGTT TTTCCGGGTT CCGGCTGGCG
    601  CTGCTGTGCG GCGGCGAGCG TTATGCCACG GTGCAAGTGG AAATTATACA
    651  GTTGTTCATG TTGGAATCTG ATATGGCGGT TGCTTCGGTG CTGTTGTGGC
    701  TGGTGTGGG GGTAAACGGG GCGGCGAGGT TGCTGTATGC GTGGTTGCGG
    751  AGGCGGCGCG TTTGCGATAA GCGCGTTTCC CCGTGTATGC CGTCCGCGCG
    801  CCGATCGGTC GGGGAATATG TGCTGCTGGC GTTGTGCGCG GCGGTGTTGT
    851  CTGTGTGCTG CCGTGTTCCT TTGTGGCAA TTGTGTGCAA AGCGTGTGTC
    901  GCGGCGGAAT CGTGCGCTGT GTTAAATGAA AGTGAACGTG GCGAGCGCGT
    951  GTGGAATACT TTGCGCTTCT CCGCGCGCGC GGTGTATGCG GCGGCGGTTT
   1001  TGCGCTGGGT GTATGCGCGG GCGGCGCGCG GGTGCGCGGT GATCGCGCGG
   1051  CTGATGTTT TGCGCTTTAT GGTGTGCGCG GTTGTGTTGT CCGCGCGCGT
   1101  CTGCTGCTT TATCGCAGT GCGCGGCTTC GTTGGCGGTG CTGCTGCGCA
   1151  TGTATGCGCT GCTGCGGTAT CCGTTTGTGG CAAGAATGT TTTATCAGCC
   1201  TGGGATGCAC TGCGCGCGGA TTACGGCAGG GCGGCGCGCG GTTGTGCGTG
   1251  AAACGCGCTT CAGACGGCAT GCGCGCATCG GTTCCCCCTC TTGAACCGCG
   1301  CGTTCGCGCG CGTCTGACTT TTGCGCGCGG CAACCTGGGT GGGCGAATTT
   1351  GCGGCGACAT TGTTTCTGTC CGGCTCGGAA TGGCAGACCG TGACGACTTT
   1401  GATTTATGCC TATTTGGGAC GCGCGGCTGA GGTATATTAC GCGCGGCGGA
   1451  TGGTCTGAC ATTGCTGTTG GCGGCGGTTC CGCTGGGTAT TTTCTGCTG
   1501  TTGGACGCGG GCGAAGCGG AAAACGACG GAAACGTTAT AA
  
```

This corresponds to the amino acid sequence <SEQ ID 576; ORF139-1>:

```

55      1  MDGRRVVVVG AFALLPSAFL AVVVVAPLMA VAAVDGLAWR AVLSDAYMLK
     51  RLAWTVFQAA ATCVLVLPLG VPVAVVLARL AFGPARVLIR LLMLPFVMPF
    101  LVAGVGVLLA FGADGLLWRG RQDTPYLLLY GNVFENLPVL VRAAYGVGVQ
    151  VPAARLQTAI TLGAGAWRRF WDIEMVPLRP WLAGGVCLVF LYCFSGGLGA
  
```

-330-

201 LLLGGSRYAT VEVEIYQLVM FELDMAVASV LVWLVLGVT AAGLLYAWFG
 251 RRAVSDKAIV FVMFSPQSV GEYVLLAFAA AVLSVCCLEF LLATVVKAWF
 301 AGESWRVLM SETWQAVWNT LRFSAAYTA RAVLGVVVFA RARRSAMMRG
 351 LMFLEWVSP VCVSGVLLV YPQWASLPI YQWYALLAY EPYAKDWLSA
 401 WDALFPDYGR AAAGLGANGF QTACRITFPL LKPALRGLT LAARTCVGEF
 451 AATFLSRPE WQTLTLLIYA YLGRAGEDNY ARAMVTLILL AAFALCIFLL
 501 LDGEGGKQT ETI*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

10 ORF139 shows 94.7% identity over a 189aa overlap with an ORF (ORF139a) from strain A of *N. meningitidis*:

```

                                10          20          30
orfl39.pep                AWSAGESWRVLMSESETWHAVWNTLRFSAAA
15 orfl39a                QSVGEYVLLAFAAAVXSVCCLEKLLATVVKAWSAGESWRVLMSESETWQAVWNTXRFSA
                                280      290      300      310      320
                                40          50          60          70          80          90
orfl39.pep                VYAAAVLGVVYAAARRSAMMRGLMFKPFMVSPVCVSAGVLLLPQWFTASLPLLLAMAYAL
20 orfl39a                VYAAAVLGVVYAAARRSAMMRGLMFLPFMVSPVCVSAGVLLLPQWFTASLPLLLAMAYAL
                                330      340      350      360      370      380
                                100        110        120        130        140        150
25 orfl39.pep                LAYPFVAKDVLSDALPPDYGRAAGLGANGFQTACRITFPLKPALRRGLTAAATCV
orfl39a                LAYPFVAKDVLSDALPPDYGRAAGLGANGFQTACRITFPLKPALRRGLTAAATCV
                                390      400      410      420      430      440
                                160        170        180        189
30 orfl39.pep                GEFAATFLFSRPEWQTLTLLIYAYLGRAGEDNYARAMVL
orfl39a                GEFAATLFXSRXEWQTLTLLIYAYXGRAGXDNVARAMVTLTLLAAAFALGXFLLLDGGEGG
                                450      460      470      480      490      500

```

35 The complete length ORF139a nucleotide sequence <SEQ ID 577> is:

```

1  ATGGATGGAC  GGCCTTGGGC  GGTATGGGGT  GCTTTGGCCC  TGCTGCCTTC
51  GGCTTTTTTG  GCGGCAATGG  TCGTTCGCC  TTTGTGGCG  GTGGCGCGCT
101  ATGACGGTTT  GCGCTGGCGC  GCGGTGCTG  CGGATGCTA  TATGCTCAA
151  CGTTTGGCGT  GGACGGTATT  TCAGGCAGC  GCAACCTGT  TGTGTGGCT
201  GCCTTTGGGC  GTGCTGTGCG  CGTGGGTGT  GCGCGGGCT  GCGTTCCGG
251  GCGGGGCTTT  GGTGCTGCGC  CTGCTGATG  TGCTTTTGT  GATGCCACAG
301  TTGTGTGGCG  GCGTGGGCGT  GCTGGCTCT  TTCGGGGCG  ACGGCTGTN
351  GTGGCGCGCG  TGCGAGGATA  CGCCGTATC  GTTGTGTAC  GGCAATGTG
401  TTTTINACCT  TCCTGTGTGT  CTCAGGCGG  CATATCAGG  GTTGTGCAA
451  GTGCCGTGCG  CACGGCTTCA  GACGCGCAC  ACATTCAGG  CGGGGCGGT
501  GCGGCGGTTT  TGGACATATT  AATGCCGT  TTTGCGCCG  TGCTTGGCG
551  GCGGCGGTGT  CTTGTCTTC  CTGATTTGT  TTTGCGGTT  CGGGCTGCCA
601  TTGCTGTGG  GCGGACGCG  TTATGCCAG  TCGTTCGGT  CTNGTGTGG
651  GTTGGCTGT  TTGCAACTCG  ATATGGCGT  TCGTGTATG  GTGTGTGGG
701  TGTGTGGG  GGTACACGCG  GCGTGTATG  CCGTGTATG  CCGCGCGCG
751  AGGCGCGCG  TTTCCGATA  GCNCTTTTC  CCGTGTATG  CCGCGCGCG
801  CGAGTCTGTC  GGGGAATAT  TGCTNCTGC  GTTTCGGCG  CGCGTGTGT
851  CTGTGTGCTG  CCGTTTTCN  TTGTTGGCA  TTGTTGTGA  AGCGTGTGT
901  GCGGCGGAAT  CGTGGCGGT  GTTATGGAA  AGTGAACGT  GCGGCGGCT
951  GTGGAATACT  NTGCGCTTCT  CGCGGCGCG  GGTGTATGC  GCGGCGGTT
1001  TGGGTGTGTT  GTATGCGCG  GCGGCGCGG  GGTGCGGCT  GATGCGGCG
1051  CTGATGTTT  TGCGCTTAT  GTGTGCGCG  GTTGTGTTT  CGCGGCGGT
1101  GCTGCTGCTT  NATCGCAGT  GGACGGCTT  GTTTCGGCT  CTGCTGCGA
1151  TGTATGGCT  GCTGGCGTAT  CCGTTTGTG  CAAAGATGT  TTTATCAGC
1201  TGNATGCAC  TGCGCGCGA  TTACGCGAG  GCGGCGGCG  GTTGGGTGT
1251  AAACGGCTT  CAGACGGCAT  GCGCATCAC  GTTCCCTCT  TTGAAACCG
1301  CTTGCGCGG  CGGTCTGACT  TTGCGCGCG  CAACCTGCT  GGGCGAATT
1351  GCGGCAACCT  TGTTCTGTCT  GCGTCNCGAT  TGGCAGACG  TGACGACTTT

```

1401 GATTATATGCC TATNTGGGAC GCGCGGGTGA NGATAATAG GCGCGGGCGA
 1451 TGGTCTGAC ATTGCTGTG GCGCGGTG CCGTGGGTAT TTCTCTGCTG
 1501 TTGACGGCG GCGAGGCGG AAGACGACG GAACGTAT AA

This encodes a protein having amino acid sequence <SEQ ID 578>:

5 1 MDGRRAWVWG AFALLPSAFL AAMVVAFLWA VAAVDGLAWR AVLSDAYMLK
 51 RLAWTVFQAA ATCVLVPLG VPVAVWLARL AFGRAVLRL LLMLPFVMT
 101 LVAGVGVLLA FGADGLXWRG WQDTPYLLLY GNVFFXLPVL VRAAYQGFVQ
 151 VPAARLOTAX TLGAGAWRRF WDIEMPVLRP WLAGGVCLVF LYCFSGFGLA
 201 LLLGGSRYAT VEVEIYQLVM FELDMVASV LVWLXVGTA AAGLLYAWF
 251 RRAVSDKAVS PVMPSPQSV GEYVLLAFAA AVXSVCCLFX LLAIIVKAW
 301 AGESWRVMESE TWQAVVWNT XRFSAAYVAA AAVLGVVYAA AARRSAMWRG
 351 LMFLLPMVSP VCVSAGVLLL XPQWASLPL LLAMYALLAY PFVAKDVL
 401 XDALPPDYGR AAGLGANGF QTACRITFPL LKPALRGLT LAATCVGEF
 451 AATLFXSRXE WQTLTLIYA YXGRAGXDNY ARAMVLTLL AAFALGXFL
 151 501 LDGEGGKRT ETL*

ORF139a and ORF139-1 show 96.5% homology over a 514aa overlap:

orf139a.pep MDGRRAWVWGAFFALLPSAFLAAMVVAFLWAAVVDGLAWRAVLSDAYMLKRLAWTVFQAA
 20 orf139-1 MDGRRAWVWGAFFALLPSAFLAAMVVAFLWAAVVDGLAWRAVLSDAYMLKRLAWTVFQAA
 orf139a.pep ATCVLVPLGVPVAVWLARLAFGRALVRLMLPFVMTPLVAGVGVLLAFLGADGLXWRG
 20 orf139-1 ATCVLVPLGVPVAVWLARLAFGRALVRLMLPFVMTPLVAGVGVLLAFLGADGLXWRG
 25 orf139a.pep WQDTPYLLLYGNVFFXLPVLVRAAYQGFVQVPAARLOTAXTLGAGAWRRFWDIEMPVLRP
 orf139-1 RQDTPYLLLYGNVFFNLPVLVRAAYQGFVQVPAARLOTARTLGAGAWRRFWDIEMPVLRP
 30 orf139a.pep WLAGGVCLVFLYCFSGFGLALLLGGGSRYATVEVEIYQLVMFELDMVASVVLWLXVGTA
 orf139-1 WLAGGVCLVFLYCFSGFGLALLLGGGSRYATVEVEIYQLVMFELDMVASVVLWLXVGTA
 35 orf139a.pep AAGLLYAWFGRRAVSDKAVS PVMPSPQSVGEYVLLAFAAAVXSVCCLFXLLAIIVKAW
 orf139-1 AAGLLYAWFGRRAVSDKAVS PVMPSPQSVGEYVLLAFAAAVXSVCCLFXLLAIIVKAW
 40 orf139a.pep AGESWRVMESE TWQAVVWNTXRFSAAYVAAAVLGVVYAAARRSAMWRGLMFLLPMVSP
 orf139-1 AGESWRVMESE TWQAVVWNTLRFSAAYVAAAVLGVVYAAARRSAMWRGLMFLLPMVSP
 45 orf139a.pep VCVSAGVLLLXPQWASLPLLLAMYALLAYPFVAKDVLXAXDALPPDYGRRAAGLGANGF
 orf139-1 VCVSAGVLLLXPQWASLPLLLAMYALLAYPFVAKDVLXAWDALPPDYGRRAAGLGANGF
 50 orf139a.pep QTACRITFLLKPALRGLTLAAATCVGEFAATLFXSRXEWQTLTLIYAYXGRAGXDNY
 orf139-1 QTACRITFLLKPALRGLTLAAATCVGEFAATLFLSRPEWQTLTLIYAYLGRAGEDNY
 55 orf139a.pep ARAMVLTLLAALFALGXFLLLDGGEGGKRTETLX
 orf139-1 ARAMVLTLLAALFALGXFLLLDGGEGGKRTETLX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF139 shows 95.2% identity over a 189aa overlap with a predicted ORF (ORF139ng) from

55 *N.gonorrhoeae*:

orf139.pep AAGSAGESWRVMESE THAVVWNTLRFSA 30
 orf139ng QSVGEYVLLAFSAVLSVCCLFPLSAIVKAWSAGESWRVMESE TWQAVVWNTLRFSA 327
 60 orf139.pep VYAAAVLGVVYAAAPARRSAMWRGLMFPMVSPVCVSAGVLLLYPQWASLPLLLIAYAL 90
 orf139ng VFYAAVVGYYAAARLLVMWRGLVFLPMVSPVCVSAGVLLLYPQWASLPLLLIAYAL 387

	orf139.pep	LAYPFVAKDVL\$AWDALPPDYGRAAAGLGANGFQTACRITFFLLKPALRRGLTAAATCV	150
	orf139.ng	LAYPFVAKDVL\$AWDALPPDYGRAAAGLGANGFQTACRITFFLLKPALRRGLTAAATCV	447
5	orf139.pep	GEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVL	189
	orf139.ng	GEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVLTLLLSAFVFCIFLLDNGEGG	507

The complete length ORF139ng nucleotide sequence <SEQ ID 579> is predicted to encode a protein having amino acid sequence <SEQ ID 580>:

10	1	MDGRCAVVRG	AFSLLPSAFL	AVMVVAPLWA	VAAVDGLAWR	AVLSDAYMLK
	51	RLAWTVFQRA	ATCVLVLPLG	VPVAVVILARL	AFPGRALVLR	LIMLPFVMP
	101	LVAGVGLAL	FGADGLLWRG	RQDTPYLLLY	GNVFNFLVL	VRAAVQGFAG
	151	VPAARLQATAR	TLGAGAWRFF	WDIEMPVLRP	WLAGGVCLVF	LYCPSGFGLA
	201	LLLGSSRYAT	VEVEIYQLVM	FELDMAGASA	LVWLVLGVTA	AAGLLYAWFG
	251	RRAVSDKAVS	PVMPSPQSV	GEYVLLAFSV	AVLSVCLLFP	LSAIVVKAWS
	301	AGESRRVIME	SETWQAVVNT	LRFSAAVFA	AAVLGVVYAA	AARRLVWMRG
	351	LVFLPFMVSP	VCVSAGVLLL	YPGWTSASLP	LLAMAYALLA	PFVAKDVL\$A
	401	WDALPPDYGR	AAAGLGANGF	QTACRITFPL	LKPALRRGLT	LAAATCVGEF
	451	AATLFLSRPE	WQTLTTLIYA	YLRAGEDNY	ARAMVLTLLL	SAFAVCIFLL
20	501	LDNGEGGKRT	ETL*			

Further work revealed a variant gonococcal DNA sequence <SEQ ID 581>:

1	ATGGATG\$AC	GGTGTGGGC	GGTACGGGGT	GCTTTTCCC	TGCTGCCCTC
51	GCCTTTTGT	ATCVLVLPLG	VPVAVVILARL	AFPGRALVLR	LIMLPFVMP
101	ATGACGGT	GGTGTGGGC	GGTGTGGGC	GGTGTGGGC	GGTGTGGGC
151	CGTTGGCGT	GGACGGTGT	TACAGCGGCG	GCAACCTGTG	TGCTGGCTGT
201	GCCTTTGGGC	GGTGTGGGC	GGTGTGGGC	GGTGTGGGC	GGTGTGGGC
251	GGCGGGCTT	GGTGTGGGC	GGTGTGGGC	GGTGTGGGC	GGTGTGGGC
301	CTGTGTGGGC	GGTGTGGGC	GGTGTGGGC	GGTGTGGGC	GGTGTGGGC
351	GTGGCGGGC	GGCGAGGATA	CGCGGTATCT	GTTGTGTGAC	GGCAATGTGT
401	TTTTCAACCT	CGCGGTGTGT	GTCAGGGGCG	CGTATCAGGG	GTTGTGTCAA
451	GTGGCTGGGC	CACGGCTTCA	GACGGCAGCG	ACGTTGGGCG	CGGGGGCGTG
501	GGCGGGCTT	TGGGACATTT	AAATGCCGTG	TTTGGCGCG	TGGCTTGGCG
551	GGCGCGGTG	CCTTGTCTTC	CTGTATTGTT	TTTGGGGT	CGGGCTGGCA
601	TTGTGTGTG	GGCGAGCGCG	TTATGCCACG	GTGGAAGTGG	AAATTTACCA
651	GTGTGTGTG	TTTGGACCTG	ATATGGCGGG	GGCTTCGGCG	CTGTGTGGCG
701	TGTGTGTGG	GGTAAACGGCG	GGCGCAGGTT	TGCTGTATGC	TGTGTTCGGC
751	AGGCGCGCG	TTTCGGATAA	GGCGGTTTCC	CCGCTGATGC	CGTCCCGCG
801	GCAATCGGT	GGGGAATATG	TATGTCTGGC	ATTTTCGGTG	GGCGGTTTGT
851	CGTGTGTGT	CCTGTTTCT	TTTGGCGCA	TTGTGTGAGA	AGCGGTGGCT
901	GGCGGCAAT	CGCGGGTGT	GTTAAATGAT	GTCAGGAGTA	GTCAGGAGTA
951	GTGGAATAC	TTGGCGCTTT	CGCGCGGCG	GGTGTGTGAG	CGCGCGGTTT
1001	TGGGTGTGT	GTATGCGCGC	GGCGCGGCG	GGCTGGTGTG	GATGCGCGGA
1051	CTGTGTGTT	TACCGTTTAT	GGTGTGCGCG	GTTGTGTGTT	CGCGGGCGGT
1101	GCTGTGCTT	TATCCGGGCT	GGACGGCTTC	GTTTACCGGT	CTGTGGCGCA
1151	TGTATGCGCT	GCTGGCGTAT	CGTGTGTG	CAAAAGATGT	TTTATCGGCC
1201	TGGGATGCAC	TGCGCGCGGA	TTACGGCAGG	GGCGGGCAG	GTTTGGGCGC
1251	AAACGGCTT	CAGACGGCAT	CGCGATACAC	GTTCGCCCTC	TTGAAACCGG
1301	CGTGTGCGCG	CGGTCTGACT	TTGGCGGCGG	CGACGTGTGT	GGGCGAATTT
1351	GGCGCAACCT	TGTTCTGTGC	GGCTTCGGAA	TGGCAGACGT	TGACGACTTT
1401	GATTTATGCC	TATTTGGGCG	GTGCGGGTGA	GGACATTTAT	GGCGGGGCAA
1451	TGTGTGTGAC	ATTGCTGTTG	TGCGCATTTG	CGGTGTGCAT	TTTCTGTGCT
1501	TTGGACAACG	GCGAAGGCGg	aaaACGGACG	GAAACGTTAT	AA

This corresponds to the amino acid sequence <SEQ ID 582; ORF139ng-I>:

1	MDGRCAVVRG	AFSLLPSAFL	AVMVVAPLWA	VAAVDGLAWR	AVLSDAYMLK
51	RLAWTVFQRA	ATCVLVLPLG	VPVAVVILARL	AFPGRALVLR	LIMLPFVMP
101	LVAGVGLAL	FGADGLLWRG	RQDTPYLLLY	GNVFNFLVL	VRAAVQGFAG
151	VPAARLQATAR	TLGAGAWRFF	WDIEMPVLRP	WLAGGVCLVF	LYCPSGFGLA
201	LLLGSSRYAT	VEVEIYQLVM	FELDMAGASA	LVWLVLGVTA	AAGLLYAWFG
251	RRAVSDKAVS	PVMPSPQSV	GEYVLLAFSV	AVLSVCLLFP	LSAIVVKAWS
301	AGESRRVIME	SETWQAVVNT	LRFSAAVFA	AAVLGVVYAA	AARRLVWMRG
351	LVFLPFMVSP	VCVSAGVLLL	YPGWTSASLP	LLAMAYALLA	PFVAKDVL\$A
401	WDALPPDYGR	AAAGLGANGF	QTACRITFPL	LKPALRRGLT	LAAATCVGEF
451	AATLFLSRPE	WQTLTTLIYA	YLRAGEDNY	ARAMVLTLLL	SAFAVCIFLL
501	LDNGEGGKRT	ETL*			

ORF139ng-1 and ORF139-1 show 95.9% identity over 513aa overlap:

	orf139ng	MDGRCAVARGAFSLLESAFLAVMVVAPLWAVAAVDGLAWRAVLSDAYMLKRLAWTVFCQA
5	orf139-1	MDGRRVWVGAFALLLESAFLAVMVVAPLWAVAAVDGLAWRAVLSDAYMLKRLAWTVFCQA
	orf139ng	ATCVLVLPGLGVFVAVVLARLAFPGRALVLRLLMLPFVMPMTLVAGVGVVLAIFGADGLLRWG
	orf139-1	ATCVLVLPGLGVFVAVVLARLAFPGRALVLRLLMLPFVMPMTLVAGVGVVLAIFGADGLLRWG
10	orf139ng	RQDTFYLILLYGNVFNLPVLVRAAYQGFVQVPAARLQTARTLGAGAWRRFWDIEMFVLRF
	orf139-1	RQDTFYLILLYGNVFNLPVLVRAAYQGFVQVPAARLQTARTLGAGAWRRFWDIEMFVLRF
15	orf139ng	WLAGGVCLVFLYCFSGFGLALLGGSRVATVEVEIYQLVMFELDMAGASALVWLVLGVT
	orf139-1	WLAGGVCLVFLYCFSGFGLALLGGSRVATVEVEIYQLVMFELDMAGASALVWLVLGVT
	orf139ng	AAGLLYAMFGRRAVSDKAVSPVMPSPQSVGEYVLLAFSAVLSVCCLPFLSAIVVKAWS
20	orf139-1	AAGLLYAMFGRRAVSDKAVSPVMPSPQSVGEYVLLAFSAVLSVCCLPFLSAIVVKAWS
	orf139ng	AGESRRLVMESETWQAVWNTLRFSAAAVFAAAVLGVVYAAAARLVMRGLVFLPFMVSP
25	orf139-1	AGESRRLVMESETWQAVWNTLRFSAAAVFAAAVLGVVYAAAARLVMRGLVFLPFMVSP
	orf139ng	VCVSAGVLLLYPGWTASLPLLLAMAYALLAYPEVAKDVLSDAWDALPDPYGRAAAGLGANGF
	orf139-1	VCVSAGVLLLYPGWTASLPLLLAMAYALLAYPEVAKDVLSDAWDALPDPYGRAAAGLGANGF
30	orf139ng	QTACRITFPLKPAIRRLGLTAAATCGVEFAATLFLSRPEWQTLTLLIYAYLGRAGEDNY
	orf139-1	QTACRITFPLKPAIRRLGLTAAATCGVEFAATLFLSRPEWQTLTLLIYAYLGRAGEDNY
35	orf139ng	ARAMVLTLLLSAFVAVCI FLLLDNGEGGKRTETL
	orf139-1	ARAMVLTLLLSAFVAVCI FLLLDNGEGGKRTETL

Based on the presence of a predicted binding-protein-dependent transport systems inner membrane component signature (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or

40 diagnostics, or for raising antibodies.

Example 70

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 583>:

	1	ATGGACGGCT	GGACACAGAC	GCTGTCCGCG	CAAACCCGT	TGGGCATTTC
45	51	GGCGCGGGCA	ATCATCTCTCA	TTCTGATTTT	AATCGTCAGA	TTCGCGATCC
	101	ACGCGCTGCT	GACACTGGTC	ATCGTCAGCC	TGCTGACGGC	TTTGCGCAACC
	151	GGTTTGCCCA	CAGGCAGCAT	TGTCACAGAC	ATACTGGTCA	AAAACCTCGG
	201	CGCACAGCTC	GGCGCGTGG	CGCTTCTGCT	CGGCTGTGGC	CGGATGCTCG
	251	AACGTTTGGT	C...			

This corresponds to the amino acid sequence <SEQ ID 584; ORF140>:

50	1	MDGWTQLSA	QTLGISAAA	IILILILVR	FRHALLPLV	IVSLLTALAT
	51	GLPTGSIVKD	ILVKNFGTL	GGVALLVGL	AMLERIV..	

Further work revealed the complete nucleotide sequence <SEQ ID 585>:

	1	ATGGACGGCT	GGACACAGAC	GCTGTCCGCG	CAAACCCGT	TGGGCATTTC
55	51	GGCGCGGGCA	ATCATCTCTCA	TTCTGATTTT	AATCGTCAAA	TTCGCGATCC
	101	ACGCGCTGCT	GACACTGGTC	ATCGTCAGCC	TGCTGACGGC	TTTGCGCAACC
	151	GGTTTGCCCA	CAGGCAGCAT	TGTCACAGAC	ATACTGGTCA	AAAACCTCGG

201 CGGCACGCTC GGGCGGTGG CGCTTCTGGT CGGCTCTGGC GCGATGCTCG
 251 GACGTTTGGT CGAAGCATCC GGGCGGCGAC AGTCGCTGGC GGCAGCGCTG
 301 ATCGGATGT TCGGCGAATA ACGGCGGCGG TTGCGGCTGG GCGTGTGCTG
 351 GCTGATTTCG GCGTTCGCGA TTTTCTCGA TGCCGGACCTA ATGCTCATGC
 401 TGCCCATCGT GTTCGCCACC GCAACGGGCA TGAACACGGA CGTACTCGCC
 451 TTGCGGCTTG CCTCCATCGG GCGATTTCCT GTCATGCACG TCTTCTCGGC
 501 GCCCATCTCG GCGCGGATGG CGGCTTCGGA ATTTTACGGC GCGAATCATG
 551 GCCAAGTTTT GATTTTGGGT CTGCGGACCG CTTTCATCAC ATGGTATTTC
 601 AGCGGCTATA TGCTCGGCAA AGTGTGGGG CGCACCATCT ATGTTCCCGT
 651 TCCCGAATCT CTCAGCGGCG GCAACGCAAGA CAACGACCTG CCGAAGAAGC
 701 CTGCCAAGAC AGGAACGGTC GTCCGCTATCA TGCTGATTTC CATGCTGCTG
 751 ATTTTCTCGA ATACCGCGGT ATCGGCGCTC ATCAGCGAAA AACTCGTAAG
 801 TGCGGACGAA ACCTGGGTTT AGACGCGCAA AATAATCGGT TCGACACCGA
 851 TCGCCCTTCT GATTTCGGTA TTGCTCGCAC TGTTTCTCTT GGGACGCAAA
 901 CGCGGCGAAA GCGGCGACCG GTTGGAAAAA ACCGTGGACG GCGCACTCGC
 951 CCCCCTCTGT TCGGTGATTTC TGATTACCGG CGCGGCGGCT ATGTTGCGGG
 1001 GCGTTTTGCG CGCTTCCGGC ATCGGCAAGG CACTCGCGCA CAGCATGGGG
 1051 GATTGGGCA TTCCGCTCCT TTTCGGTGTG TCTCTGTGTC CTTTGGCACT
 1101 GGTATCGCG CAAGGTTTGG CAACCTCTGC CTTGACACAC GCGCGCGGCG
 1151 TGATGCTGCC TGCCGTTGCC GCGCGCGGCT TTACGCACTG CGAGCTGGCC
 1201 TGTATCGTAT TGSCAACGGC GCGAGGTTGG GTGCGTTGCA GCCACTTCAA
 1251 CGACTCGCGG TTCTGGCTGG TCGGCGCTCT CTTGGACATG GACGTACCGA
 1301 CCAAGCTGAA AACCTGGACG GTCAACCGAA CCTCATCGC ACTCATCGGG
 1351 TTTGCTTGT CCGCACTGCT GTTCGCCATC GTCTGA

25 This corresponds to the amino acid sequence <SEQ ID 586; ORF140-1>:

1 MDGWTQTLA QTLGISAAA IILILILIVK FRIHALTLV IVSLLTALAT
 51 GLPTGSIVND ILVKNFGGTL GSVALLVGLG AMLGRVLVTS GGAQSLADAL
 101 IRMFGEKRAP FALGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVLP
 151 FALASIGAFS VMHVFLPPH GPAAASEFYG ANIGQVLLG LPTAFITWYF
 201 SGYMLGKVLG RTIHRVPEL LSGGTQDNDL PKEPKAGATV VAMLIPLML
 251 IFLNTGVSAL ISEKLVSADE TWVQAKIIG STPIALLISV LVALFVLGRK
 301 RGESSSALEK TVDGALAPVC SVILITGAGG MFGGVLRAAG IKGALADMSA
 351 DLGIPVLLGC FLVALALRIA QGSATVALT AALMAPAVA AAGFTDWOLA
 401 CIVLATAGS VGSHFNDSG FVLVGRLLDM DVPTTLTKTW VNGTLIALIG
 451 FALSALLFAI V*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF140 shows 95.4% identity over a 87aa overlap with an ORF (ORF140a) from strain A of *N.*

meningitidis:

40 10 20 30 40 50 60
 orf140.pep MDGWTQTLAQTLLGISAAAIIILILILIVRFRIHALTLVIVSLLTALATGLPTGSIVKD
 orf140a MDGWTQTLAQTLLGISAAAIIILILILIVKFRHALTLVIVSLLTALATGLPTGSIVND
 45 70 80
 orf140.pep ILVKNFGGTLGGVALLVGLGAMLERLV
 orf140a VLVKNFGGTLGGVALLVGLGAMLERLVTSGGGAQSLADALIRMFGEKRAPFALGVASLIF
 50 70 80 90 100 110 120

The complete length ORF140a nucleotide sequence <SEQ ID 587> is:

1 ATGGACGGCT GGACACAGAC GCTGTCGCGC CAACCCCTGT TGGGCATTTC
 51 GGGCGGCGCA ATCATCTCTCA TTCTGATTTT AATCGTCAA TTCCGCATCC
 101 ACGCGCTGCT GACACTGGTC ATGCTCAGCC TGCTGACGGC TTTGCGAAC
 151 GGTTTGCCCA CAGGCAGCAT GTCTCAACGAC GTACTGCTCA AAAACTTCGG
 201 CGGCACGCTC GGGCGGCTGG CGCTTCTGGT CGGCTCTGGC GCGATGCTCG
 251 GACGTTTGGT CGAAGCATCC GGGCGGCGAC AGTCGCTGGC GGCAGCGCTG
 301 ATCGGATGTT TCGGCGAAAA ACGCGCACCG TTGCGGCTGG GCGTGTGCTC
 351 GCTGATTTCG GGCTTCCCGA TTTTCTCGA TGCCGGACCTA ATGCTCATGC
 401 TGCCCATCGT GTTCGCCACC GCAACGGGCA TGAACACGGA CGTACTCGCC
 451 TTGCGGCTTG CCTCCATCGG GCGATTTCCT GTCATGCACG TCTTCTCGGC

501 GCCCATCCG GCGCCGATTG CCGCTTCCGA ATTTTACGGC GCGAACATCG
 551 GCGAAGTTT GATTTTGGGT CTGCCAGCGC CTTTCATTCG ATGGTATTTT
 601 AGCGGTCTATA TGCTCGGCAA AGTGTGGGGC CGCACCATCC ATGTTCCCGT
 651 TCCGCAACTG CTCAGCGCGG CGACGCAAGA CAACGACCTG CCGAAGAAGC
 701 CTGCCAAGC AGGAACGGTC GTCCGCATCA TGCTGATTCC CATGCTGCTG
 751 ATTTTCTCTGA ATACCGGCGT ATCGGCCCTC ATCAGCGAAA AACTCGTAGA
 801 TGCGGACGAA ACCTGGGTTT AGACGCGAAA AATAATCGGT TCGACACCGA
 851 TCGCCCTTCT GATTTCCGTA TTGGTGGCAC TGTTTGTCTT GGGACGCAAA
 901 CGCGGCGAAA CGCGGACGGC GTTGGAAGAA ACCGTGGAGC GCGCACTCGC
 951 CCGCGTCTGT TCCGTGATTC TGATTACCGC CGCGGGCGGT ATGTTGGCGG
 1001 GCGTTTTCGC CGCTTCCGGC ATCGGCAAGC CACTCGCCGA CAGCATGGCG
 1051 GATTTGGGCA TTCCCGTCTT TTTGGGCTGT TTCTTGTGCG CTTGGCACT
 1101 GCGTATCGCG CAAGGTTTCGG CAACCGTCGC CCGTACCACC CGCGCGCGCG
 1151 TGATGGCTCC TGCGGTTGCC CGCGCGGCGT TTACCGACTG GCAGCTCGCC
 1201 TGATCGTAT TGCGCAACGGC GCGAGGTTGC TGCGGTGCGA GCGCATTCGA
 1251 CGACTCGGCG TTCTGGCTGG TGGCGCGCTT CTTGGACATG GACGTACCGA
 1301 CCACGCTGAA AACCTGGAGC GTCAACCAAA CCTCATCGCG ACTCATCGCG
 1351 TTTGCTTGT CCGCATGCT GTTCGCCATC GCTCA

This encodes a protein having amino acid sequence <SEQ ID 588>:

20 1 MDGWQTLSA QTLGISAAA IILILILIVK FRIHALTLIV IVSLLTALAT
 51 GLFTGSIVND VLVKNFGTLL GGVALVLGLG AMLGRLVETS GGAQSLADAL
 101 IRMFGEKRAP FALGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVLE
 151 FALASIGAFS VMHVFLLPHF GPIAASEFYG ANIGQVLILG LPTAFITWYF
 201 SGYMLGKVLG RTIHVPPEL LSGGTQDNDL PKPEPAKAGTV VAIMLIPMLL
 251 IFIANGVSAI ISEKLVSADE TWQTAKIIG STPIALLISV LVALFVLGRK
 301 RGESESALEK TVDGALAPVC SVILITGAGG MFGGVLRAAG IGKALADMSA
 351 DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAFAVA AAGFTDWOLA
 401 CIVLATAAGS VGCSEHNSDG FWLVGRLLDM DVPTTLKTWT VNOTLIALIG
 451 FALSALLFAI V*

30 ORF140a and ORF140-1 show 99.8% identity over a 461aa overlap:

orf140-1.pep	MDGWQTLSAQTLGISAAAIIILILILIVKFRHALTLIVISLLTALATGLPTGSIVND	60
orf140a	MDGWQTLSAQTLGISAAAIIILILILIVKFRHALTLIVISLLTALATGLPTGSIVND	60
orf140-1.pep	ILVKNFGTLLGGVALVLGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFALGVASLIF	120
orf140a	VLVKNFGTLLGGVALVLGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFALGVASLIF	120
orf140-1.pep	GFPIFFDAGLIVMLPIVFATARRMKQDVLPFALASIGAFSVMHVFLLPHGPPIAASEFYG	180
orf140a	GFPIFFDAGLIVMLPIVFATARRMKQDVLPFALASIGAFSVMHVFLLPHGPPIAASEFYG	180
orf140-1.pep	ANIGQVLILGLPTAFITWYFSGYMLGKVLGRTIHVPPELLSGGTQDNDLPKPEPAKAGTV	240
orf140a	ANIGQVLILGLPTAFITWYFSGYMLGKVLGRTIHVPPELLSGGTQDNDLPKPEPAKAGTV	240
orf140-1.pep	VAIMLIPMLLIFLNTGVSALISEKLVSADETWWQTAKIIGSTPIALLISVLVALFVLGRK	300
orf140a	VAIMLIPMLLIFLNTGVSALISEKLVSADETWWQTAKIIGSTPIALLISVLVALFVLGRK	300
orf140-1.pep	RGESESALEKTVDGALAPVCSVILITGAGGMFGGVLRAAGIGKALADMSADLGIPLVLLGC	360
orf140a	RGESESALEKTVDGALAPVCSVILITGAGGMFGGVLRAAGIGKALADMSADLGIPLVLLGC	360
orf140-1.pep	FLVALALRIAQGSATVALTTAAALMAFAVAAGFTDWOLACIVLATAAGSVGCSEHNSDG	420
orf140a	FLVALALRIAQGSATVALTTAAALMAFAVAAGFTDWOLACIVLATAAGSVGCSEHNSDG	420
orf140-1.pep	FWLVGRLLDMDVPTTLKTWTVNOTLIALIGFALSALLFAIV	461
orf140a	FWLVGRLLDMDVPTTLKTWTVNOTLIALIGFALSALLFAIV	461

Homology with a predicted ORF from *N.gonorrhoeae*

ORF140 shows 92% identity over a 87aa overlap with a predicted ORF (ORF140ng) from

65 *N.gonorrhoeae*:

	orf140.pep	MDGWQTQLSAQTLTGISAAAILLILILIVFRFIHALLLTVIVSLTALATGLPTGSIVKD	60
	orf140ng	MDGRQTLSAQTLTGISAAAILLILILIVKFRIRALLTVIASLLTALATGLPTGSIVND	60
5	orf140.pep	ILVKNFGGTLGGVALLVGLGAMLERLV ::: ::	87
	orf140ng	VLVKNFGGTLLGGVALLVGLGAMIGRLVETSSGAQSADALIRMFGKEKRAPFAGVASLIIF	120

The complete length ORF140ng nucleotide sequence <SEQ ID 589> was predicted to encode a protein having amino acid sequence <SEQ ID 590>:

```

10      1  MDGRTOTLSA  O'LKLGISRAAA  ILIILILVL  FRIRALRLTV  GASLTLTALAT
51      2  _LPTGSIIVND  VLNVKPGFTL  GOVALLVGLG  AMGLRLVETS  IAGQSLADAL
101     3  IRMFGEKRAP  FAPGVASLIF  GPTIFDFDGL  IVMLPIVTEAT  ARRMKQDVLP
151     4  FALASVAGS  VMHVFLPHP  PQAIAASEYF  ANIGQVGLIL  LPTAFTWYF
201     5  SGVMLKGVIG  RAITHVPPEL  LSGGOTQSDP  KPEKPAAGTV  VAVMLPLML
251     6  ILFTNGVSL  ISKRLYSAE  TWVGTQATSP  STPAVLSL  LAALLVLGRK
301     7  RGECSPTLKG  TVDGLALAPC  SVAILTAGG  MFGVLVLRAG  IKGALADSMA
351     8  DIGIPVLGCG  FLVALAPLA  QGSAVTALTT  AAALMAPAVA  AAGFTDQILA
401     9  CIVLATRAGS  VGCSHFNDSE  FWLVGRLSDM  DVPTTLTKWT  VNQTILAFIG
451    10  FALSAIFVA  V*

```

20 Further work revealed a variant gonococcal DNA sequence <SEO ID 591>:

		ATGGAAGGGCC	GGACACAGAC	GCTGTCCGG	CAAACTGTCT	TGGGCAATTC
	51	GGCGGGCGGCA	ATCATCCCTCA	TTCTGATTA	AATGTCGTA	TTCCGATCC
	101	GGCGGCTGCT	GACATCGTGC	ATGCGCATCC	TGCTGACGGC	TTTGCAACCC
25	151	GGTTTGGCCCA	CAGCGAGCAT	GCTCAACGAC	GTACTGTGCA	AAAACCTCGG
	201	CGCGACGCTC	GGCGGCGTGG	GGCTCTCGGT	CGGCTGTGGC	GCAATCGTGC
	251	GACGTTTGGT	AGAAACATCC	GGCGGCGATC	AGTGGCTGGG	GAGACGGCTG
	301	ATCGCGATGT	TGCGGCAAAA	ACGGCGACCG	TGCGTCCGG	CGGTTGCCTC
	351	GCTGATTTTC	GGCTTCCGGA	TTTTCTTGCA	TGCGGACTA	ATCGTATCG
	401	TGCGCATCGT	ATTGCGCCAC	CAGCGGCGCA	TGAACAGCA	GTGATCGGCC
	451	TTGCGGCTGT	CTCTCGGTGG	GCGGATTTTC	GTGATGACG	CTTCTCGGCG
	501	GGCGGCTGCT	GGCGGATGCT	GGCGGATGCT	GGCGGATGCT	GGCGGATGCT
	551	CGCAGTTGGT	GATTTTGGTG	TGCGGACGCG	CGCTCATCAC	ATGGTTATTA
	601	AGCGCGCTA	TGCTTCGGCA	AGTGTGGG	CGCGGCATC	ATGTTCCGCT
35	651	TCCGGAATCT	AGCAGCGCG	CGCGGACAGA	CAGGAGATCC	CCGGAAGAAGC
	701	CTGCCAAGC	ATGGAAGGTC	TGCGCGCTGA	TGCTGATTC	CATGCTCGTG
	751	ATTTTTCCTGA	ATACCGGGGT	ATGACCGCTG	ATCAGCGAAA	ATCATGTAAG
	801	TGCGGACGAA	ACTTGGGTAT	AGACCGCAAA	AATGATGATC	TGCGACACTG
	851	TGCGCCTTCT	GATTTCCGTA	TGCGCGCAC	TGTTTGGTCT	GGGACCGAAA
40	901	CGCGGCGAAA	GGCGGACGAC	ATCGTGAGC	CGCGGCGGAC	GGCGACCTGA
	951	CCCGCGCTGT	TCCGCTGATC	TGATTACCGG	CGCGGGCGCT	ATGTTGGGG
	1001	GGGTTTTCG	GGCTTCCGCG	ATGCGACAG	CTGTCGGA	CAGCATGACG
	1051	GATTTGGGCA	TAGCGGCTCT	TTTGGGCTGT	TTCTCTGTG	CGTTGGACAT
	1101	GGTATCCGG	TAAGGTTTCG	CACAGCTGCG	CTGACGACCA	CGCGCGGGC
45	1151	TGATCGGCTG	TGCTGCTGCG	GGCGGCGCTG	GGCGGCGCTG	GGCGGCGCTG
	1201	TGTCATCTGT	TGGACACGCG	GGCAGGTTGC	GTGGTGTGA	GCACCTACCA
	1251	CGACTCCGGC	TTCTGCTGGT	TGCGGCGCGT	CTTGGATGCG	GAGTACCGA
	1301	CACAGCTGAA	AACCTGGACG	GTCACACCA	CCCTCATCGC	ATTGATCGCG
	1351	TTTGCGTGTG	CGCGCATGCT	GTTTGCGGAT	TCGTGA	

This corresponds to the amino acid sequence <SEQ ID 592; ORF140ng-1>:

50		MDGRTQTLSA	QTLGLISAAA	IJILLILVLG	FRIRALRLTV	IASGLTALAT
	51	GLPTG3IVND	VLVKNVGGP	GOVALLVGLG	ANMLGRVLTV	GAGSLADLAE
	101	IRNMFCEKRAP	FAPFGVASLIF	PHPTPFDRLG	IVMLPTVEAT	ARNMKQDVLN
	151	FALASVGEPS	VNMHVLIF	GPPIAAEEFVG	ANMIGQVLIG	LPTAFITWYF
55	201	SGYMLKGVIG	RAIHPVPEL	LSGGDTQMD	PKPEKAGACTV	VAVMLIPMLL
	251	IFLNTVNGV	ISEKLVSDD	TWVGTAKMIS	STPALLVSI	LAALLVLGRK
	301	RCESGSTLEK	TYDGLAAPAC	SVYLITGAGG	MFGVLVRAC	IGKALDAMDS
	351	DIGIPVLLGC	FLVALALRIA	QGSATVALT	AALAMPAVA	AAGATDNLQA
	401	CIVLATRAGS	VGCSFHNDSG	FWLVGRLDM	DVPTTLTKWT	VNQTLLIAFIG
	451	ELASALLFAI	V*			

60 ORF140_{ng-1} and ORF140-1 show 96.3% identity over 461aa overlap:

orf140ng-1.pep MDGRTQFLSAQTLLGISAAAIILILILIVKFRIRALLTLVIASLLTALATGLPTGSIVND

```

    orf140-1      MDGWTQTLSAQTLGLISAAAIIILILIVKFRHALLTLVIVSLLTALATGLPTGSIVND
    orf140ng-1.pep VLVKNFGGTIGGVALLVGLGAMLGRIVETSGGAQSLADALIRMGPEKRAPFAGVASLIF
5      orf140-1      ILVKNFGGTIGGVALLVGLGAMLGRIVETSGGAQSLADALIRMGPEKRAPFAGVASLIF
    orf140ng-1.pep GFPIFFDAGLIVMLPIVFATARRMKQDVLFPALASVGAFSVMHVFLPPHGPPIAASEFYG
10      orf140-1      GFPIFFDAGLIVMLPIVFATARRMKQDVLFPALASVGAFSVMHVFLPPHGPPIAASEFYG
    orf140ng-1.pep ANIGQVLLGLPTAFITWYFSGYMLGKVLGRAIHVPFPELLSGGTQSDDPPEKPAKAGTV
    orf140-1      ANIGQVLLGLPTAFITWYFSGYMLGKVLGRTHVPFPELLSGGTQSDNLPPEKPAKAGTV
15      orf140ng-1.pep VAVMLIPMLLIFLNTGVSALISEKLVSADETWVQTAKMIGSTPVALLISVLAAALLVLGRK
    orf140-1      VAIMLIPMLLIFLNTGVSALISEKLVSADETWVQTAKIIGSTPVALLISVLVAFLVLGRK
20      orf140ng-1.pep RCGSGSTLEKTVDGALAPVCSVILITGAGMFGGVLRASGIGKALADSMADLGIPIVLLGC
    orf140-1      RCGSGSALEKTVDGALAPVCSVILITGAGMFGGVLRASGIGKALADSMADLGIPIVLLGC
    orf140ng-1.pep FLVALALRIAQSATVALTTAAALMAFAVAAAGFTDWQLACIVLATAAGSVGCSHFNDG
25      orf140-1      FLVALALRIAQSATVALTTAAALMAFAVAAAGFTDWQLACIVLATAAGSVGCSHFNDG
    orf140ng-1.pep FWLVGRLLDMVDPTTLKTWTNVNQTILAFIGFALSALLFAIV
    orf140-1      FWLVGRLLDMVDPTTLKTWTNVNQTILAFIGFALSALLFAIV

```

30 Furthermore, ORF140ng-1 is homologous to an *E.coli* protein:

```

gi|882633 [U29579] ORF_o454 [Escherichia coli] >gi|1789097 (AE000358) o454;
This 454 aa ORF is 34% identical (9 gaps) to 444 residues of an approx. 456 aa
protein GNTP_BACLI SW: P46832 [Escherichia coli] length = 454
Score = 210 bits (529), Expect = 1e-53
Identities = 130/384 (33%), Positives = 194/384 (49%), Gaps = 19/384 (4%)

Query: 88 ETSGGAQSLADALIRMGPEKRAPFAGVASLIFGFPIFFDAGLIVMLPIVFATARRMKQD 147
E SGA+SLA+ R G+KR A +A+ G P+FD G I++ PI++ A+ K
Sbjct: 80 EHSGAESLANVFSRKLGDKRTIAALTAAFFLGIPVFVDVGFIIAIIYGFQAKVAKIS 139

Query: 148 VLPFALASVGAFSVMHVFLPPHGPPIAASEFYGANIGQVLLGLPTAFITWYFSGYMLGK 207
L F L G +HV +PPHGF+AA+ A+IG + I+G+ + I GY K
Sbjct: 140 PLKFGPLPVAGIMLTIVHVAVPPHGFVAAAGLLHADIGWLTIIIGIAIS-IPVGVGVYFAAK 198

Query: 208 VLGRAIHVPFPELL-----SGGTQSDDPPEKPAKAGTVVAVMLIPMLLIFLNTGV 257
++ + + E+L G T+ SD P A V +++IP+ +I T
Sbjct: 199 IINKRQYAMSVEVLEQMQLAPASEEGATKLSDKINPPGVA-LVTSLIVIPAIIMAGT-- 255

Query: 258 SALISEKLVSADETWVQTAKMIGSTPXXXXXXXXXXXXXXXXXKRKRGESGSTLEKTVDGALA 317
+ S L+ + T ++IGS +RG S + AL
Sbjct: 256 ---VSATLMFPHSHLLGLTQLIGSPFVMAIMIALVLAFWLLALRRGWSLQHTSDIMGSALP 312

Query: 318 PACSVILITGAGMFGGVLRASGIGKALADSMADLGIPIVLLGCFVALALRIAQSGSXXXX 377
A VIL+TGAGG+FG VL SG+GKALA+ + +P+L F++LALR +QGS
Sbjct: 313 TRAVVILVTGAGGVFGKVLVESGCGKALANMLQIMDLPLLPAAFIISIALRASQGS--AT 370

Query: 378 XXXXXXXXXXXXXXXXGFTDWQLACIVLATAAGSVGCSHFNDSGFWLVGRLLDMVDPTTLK 437
G Q + LA G +G SH NDSGFV+V + L + V LK
Sbjct: 371 VALLTTGGLSEAVMGLNPIQCVLVTAAACFGGLGASHINDSGFWIVTKYGLSVADGLK 430

Query: 438 TWTNVNQTILAFIGFALSALLFAIV 461
TWTV T++ F GF ++ ++A++
Sbjct: 431 TWTVLTTLGFTGLITWCWMAVI 454

```

Based on this analysis, including the identification of the presence of a putative leader sequence

65 (double-underlined) and several putative transmembrane domains (single-underlined) in the

gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 71

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 593>:

```

5      1  ..GATTTCGGCA  TATCGCCCGT  GTATCTTTGG  GTTGCCGCCG  CGTTCAAAAC
51     TTTGCTGTGG  CGGTGGGGTG  CCGACTCATA  CGATGTGCGA  CGCTTTCGAG
101    GCGTATTTTGT  TCGCGTTATC  GCACTGACTT  CTTGCGGCTT  TGCCGGTTTC
151    AACTTTTGTG  GCACACACCA  CGCGGCAC .  GTGCTCTGGA  TTCTCATCGG
201    CTGTATCGGG  CTGATTCCAG  TTGCCCATTT  CTTCAACCCC  GCTGCGCGCG
10     251  CCTTTGCCGC  CGCCGGAGTG  GTGCTGCACT  GTTATTCTTT  GGCTGCGCCG
301    CGCGTGATTG  CGCGCTCTTT  TCTGCTCGGT  ACGGCTGGA  CGCTGATGTC
351    GTTGGCAGCA  GCTTATCCGG  CAGCATTTGC  CTTGATGCTG  CCCTTGCCCG
401    TACTGATGTT  TTTCCGTCG  ..

```

This corresponds to the amino acid sequence <SEQ ID 594; ORF141>:

```

15     1  ..DFGISPVYLV  VAAAFKHLIS  FWAADSYDVA  RFAGVFFAVI  GLTSCGFAGF
51     NFLGRHGRX  VVLLIGICIG  LIPVAHFLNP  AAAFAAAGL  VLHGYSLARR
101    RVIAASFLLG  TGMILMSLAA  AYPAAAFALM  PLPLVMFFRP ..

```

Further work revealed the complete nucleotide sequence <SEQ ID 595>:

```

20     1  ATGCTGACCT  ATACCCCGCC  CGATGCCCGC  CGGCCCGCCA  AAACCCACGA
51     AAAGCCGTGG  TGTGCTGCTG  TGATGGCGTT  TGCTGTGTTG  TGCCCGCGCG
101    TGTTTTCCCA  CGATTGTGAG  AATCTCGAGC  AACCTGCGGT  CTATACCGCG
151    GTCGAAGCAC  TGGCAGCGAC  CCCCACCCCC  TTGGTTGCCG  ATCTGTTGAG
201    TCAAAACGAT  TTCGCGATAC  CGCCCGTGTA  TCTTTGGGTT  CGCGCGCGGT
25     251    TCAAAAGATT  CTGTGCGCGC  TGGGCTGCGC  ACTCATACGA  TGCCGACAGC
301    TTTCGACGCG  TATTTTGTGG  CGTTATCGGA  CTGACTTCTG  CGCGCTTGCG
351    CGGTTTCAAC  TTTTTCGGCA  GACACACAGC  CGCAGCGTTC  GTCTGATGTC
401    TCATCGGCTG  TATCGGCGTC  ATTCCAGTTG  CCGATTTCCT  CAACCGCGCT
451    CGCGCGCGCT  TTGCGCGCGC  CGGACTGGTG  CTGACAGGTT  ATCTTTGCGC
501    TCGCGCGGCG  GTGATTGCGC  CCTCTTTTCT  GCTCGGTACG  GGCTGGACGC
30     551    TGATGTGCTT  GGCACGAGCT  TATCCGCGAG  CATTTGCCCT  GATGCTGCCG
601    TTGCCCGTAC  TGATGTTTTT  CCGTCCGTGG  CAAAGCAGGC  GTTTGATGTT
651    GAGCGGACGC  GCCTCACTTG  CCTTTGCCCT  GCGGCTTATG  ACCGTTTACC
701    CGCTGCTCTT  GGCAAAACAG  CAGCCCGGCG  TGTTCGCGCA  ATGGCTCGAC
751    TATCACGTTT  TCGGATACGT  CGGCGCGGTG  CGGCACGTTT  AGACGCGATT
35     801    CAGTTTGTGT  TACTATCTGA  AARACCTGCT  TTGTTTGA  TTGCGCGCGC
851    TGCGGCTGGC  GGTTTGGAGC  GTTTCGCGCA  CGCGCTGTTT  TTGACCGAC
901    TGGGGGATT  TGGGCGTCTG  CTGGATGCTT  GCGGTTTGG  TGCTGCTTGC
951    CGTCAATCCG  CAGCGTTTTT  AGGATAAECT  CGTCTGGGCT  CTTCCGCGCG
1001    TTGCCCTGTT  CGGCGCGGCG  CAACTGGACA  GCCTGAGGCG  CGGCGCGGCG
40     1051   GCGTTTGTCA  ACTGGTTTGG  CATTATGGCG  TTGCGAGTGT  TTGCCGTTGT
1101   CCGTGGAGCG  GCGCTTTTTC  CGATGAATTA  CGGCTGGCCC  GCCAAGCTTG
1151   CGGAACGCGC  CGCTATTTCG  ASCCGGATT  AGTTTCTGTA  TATGATCCG
1201   ATTCGATGCG  CGGTTGCCGT  ACTGTTGACA  CCTGTGPGCG  TGTGGGATG
1251   TACCGCGAAA  AACATACGCG  CGAGCGAGCG  GGTACCAAC  TGCGCGGAG
45     1301   GCGTTACCTT  GACCTGGGCT  TTGCTGATGA  CGCTGTTCTC  CGCGTGGCTG
1351   GACGCGGGGA  AAGGCCACGC  CGCGTCTGTC  CGGAGTATGG  AGGCATCGCT
1401   TTCCCGCGAA  TTGAACGGG  AGCTTTCAGA  CGGCATCGAG  TGTATCGGCA
1451   TAGCGCGCGG  CGACCTGCAC  ACGGGATTG  TTTGAGCGCA  GTACGGCACA
50     1501   TTGCGCGACC  GCGTGGCGGA  TGTCAATATG  CGCTACCGCA  TCGTCTCTCT
1551   GCCCAAAAAT  CGGGAATGCG  CGCAAGGCTG  GCAGACGGTT  TGCGAGGGTG
1601   CGGCTCCGCG  CACAAAGAC  AGTAAGTTTC  CACTGATACG  GAAAATCGGG
1651   GAAAATATAT  AA

```

This corresponds to the amino acid sequence <SEQ ID 596; ORF141-1>:

```

55     1  MLTYTPPDAR  PPAKTHEKFW  LLLLMFAFWL  WPGVFSHDLW  NPDEPAVYTA
51     VEALAGSPTP  LVAHLFGQTD  FGTPPVYLVW  AAAFKHLISF  WAADSYDAAR
101    FAGVFFAVIG  LTSCGFAGFN  FLGRHGRSRV  VLILIGICIG  LIPVAHFLNP
151    AAFAAAGGLV  LHGYSLARRR  VVIAASFLLG  TGMILMSLAA  YPAFAALMLP
201    LPVLMFFRPP  QSRRLMLTAV  ASLAFALFLM  TVPLVLIART  QPALFAQWLD

```

251 YHVEGTGGV RHVQTAFSLF YYLKNLWFA LPALPLAVWT VCTRLESTSD
 301 WGIIGVWML AVTLVAVNP RFQDNLVNL LPPLALFGAA QLDSLRKGA
 351 AFNWFEGIMA FGLFAVLWT GFPMNNGWF AKLAERAYF SPYYVPDIDP
 401 IFMAVAVLT PLWMAITRK NIRGRQAVTN WAAGVTLTWA LMTLFLPLWL
 451 DAAKSHAPVV RSMEASLSPE LKRELSDGIE CTIGGGGLH TRIVTQYGT
 501 LPHRVGDVQC RYRIVLLPQN ADAPQGWQTW WQGARPNKDK SKFALIRKGT
 551 ENI*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N. meningitidis* (strain A)

- 10 ORF141 shows 95.0% identity over a 140aa overlap with an ORF (ORF141a) from strain A of *N. meningitidis*:

```

                                10      20      30
orf141.pep      DFGISPVYLWVAARFKHLLSPWAADSYDVA
15 orf141a      WNPDEPAVYTAVEALAGSPTFLVAHLFGQIDFGIPPVYLWVAARFKHLLSPWAADFYDAA
                40      50      60      70      80      90

orf141.pep      RFAGVFFAVIGLTSCTGFAENFLGRHHGRXVVLILIGCTGLIPVAHFNLNPAFAAFAAAGL
20 orf141a      RFAGVFFAVVGLTSCGFAENFLGRHHGRSVVLILIGCTGLIPTVHFLNPAFAAFAAAGL
                100     110     120     130     140     150

                100     110     120     130     140
orf141.pep      VLGHSYLARRRVIAASFLLTGWTLMSLAAAYPAFAFALMLPLVLMFFRP
25 orf141a      VLGHSYLARRRVIAASFLLTGWTLMSLAAAYPAFAFALMLPLVLMFFRPWQSRRRLMTLA
                160     170     180     190     200     210

30 orf141a      VASLAFALPLMTVYPLLLAKTQPALFAQWLDHVFGTFGGVRHIQTAFSLFYLYLKNLWLF
                220     230     240     250     260     270

```

The complete length ORF141a nucleotide sequence <SEQ ID 597> is:

```

1  ATGCTGACCT ATACCCCGCC CGATGCCCGC CGCCCGCCCA AAACCCACGA
51 AAAGCCGTGG CTGTTGCTGT TGATGGCGTT TGCCGTGTGT TGGCCGGCGG
35 101 TGTTTTCCCA CGATTGTGTG AATCTCTGAC AACCTGCGCT CTATACCCGC
151 GTCGAAGCAC TGGCAGGCAG CCCCACCCCT TTGGTTGCCC ATCTGTTGCG
201 TCAAAATCGAT TTCGGCATAC CGCCCGTGT TCTTTGGGTT GCCGCGCGGT
251 TCAAAACATT GCTGTCCGCG TGGGCTCGCG ACCCGTATGA TGGCGCAAGC
40 301 TTTGCCGGCG TGTTTTTCGC CGTTGTGCGA CTGACTTCCT GCGGCTTTGC
351 CGGTTTCAAC TTTTGGGCGA GACACACGCG GCGCAGCGTC TCCTGATTTC
401 TCATCGGCTG TATCGGCGCT ATTCGACCGG TACACTTTCT CAACCCCGCT
451 GCGCGCGCCT TTGCGCGCGC CGGACTGGTG CTGACGCGTT ATCTTTTGGC
501 TCGCGCGCGC GTGATTGCGG CCTCTTTTCT GCTCGGTACG GGTGTGACGC
551 TGTATGTCGT GGCAGCAGCT TATCGCGCGG CATTGCGCCT GATGCTGCCG
45 601 CTGCGCGTGC TGAATGTTTT CCGTCCGTGG CAAAGCAGGC GTTTGATGTT
651 GAAGCGCAGT GCTGTGCTGT CGCTGTCCTT CGCGCTATAG ACCGTTTACC
701 CGCTGCTGCT GCAGAAAGCG CAGCCCGCGC TGTTCGGCGA ATGGCTCGAC
751 GATCAGCTTT TCGTAGCTTT CCGCGGGGTG CGGCACATTG AGAGCGCATT
801 CAGTTTGTTT TACTATCTGA AAACCTGCT TTGTTTGA TAAGCTGGCG
50 851 TGCCTGTGGC GGTTTGGACG GTTTGGCGCA CGCGCGCTGT TTGCAACGAC
901 TGGGGGAATT TGGGCGCTGT CTGGATGCTT GCGCTTTGGT TGCTGTCTGC
951 CGTCAATCGC CAGCGTTTTC AGGATAACCT CGTCTGGCTG CTTCCGCGCG
101 TTGCGCTGTT CGCGCGGGCG CAACTGGACA GCCTGAGACG CGCGCGGGCG
55 1051 GCGTTTGCA ACTGTTTCGG CATTATGGCG TCTCGACTGT TTGCGCTGTT
1101 CCGTGTGACG GCGTTTTTTC CCAATGAATTA CGGCTGGCCC GCCAAGCTTG
1151 CGGAACGGCG CGCTATTTC AGGCCGTATT ATGTTCTCGA TATCGATCCC
1201 ATTCGGATGG CGGTTGCGGT ACTGTTTACA CCCTTGTGGC TGTGGCGGAT
1251 TACCCGCAAA AACATACGCG CGAGCGGAGC GGTTACCAAC TGGCGCGCAG
1301 GCGTTACCTT GACCTGGGCT TTGCTGATGA CGCTGTTCTT CGGTGTGGCT
60 1351 GACGCGGCGA AAAGCCACGC GCGCGTGTCT CGSAGTATGG AGGCATCGCT
1401 TTCCCGGAAA TTAAGACGGG AGCTTTTACA GCGCATCGAG TGATTCGACA
1451 TAGGCGGGGG CGACCTACAC ACCTGGGATG TTTGACGCGA GTACGCGACA
1501 TTGCGCGACC GCGTCCGCGA GTACATATGC CGCTACGCGA TCGTCCGCTT
1551 GCCCAAAAC GCGGATGCGC CGCAAGGCTG GCGACGCTG TGGCAGGGTG

```

1601 CGCGCCCGCG CAACAAAGAC AGTAAGTTCG CACTGATACG GAAACCCGGG
1651 GAAATATAT TAAAAACAC AGATTGA

This encodes a protein having amino acid sequence <SEQ ID 598>:

```

1  MLTYTPPDAR PPAKTHEKFW LLLMAFAWL WPGVFSHDLW NPDEPAVYTA
5  51  VEALAGSPTP LVAHLFGQID FGIPPVYLWV AAFKHLSP WAADPYDAAR
101 FAGVEFFAVVG LITSCGFAGFN FLGRHHGRSV VLILIGCIGL IPTVHFLNPA
151 AAAFAAAGLV LHGYSLARRR VLAASFLLGT GWTLMSLAAA YPAAFALMLP
201 LPVLMFFRFW QSRRLMTAV ASLAFALPLM TVYPLLAKT QPALFAQWLD
251 DHVFGTGGV RHIOATFSLF YYLKNLLWFA LPAALPAVMT VCRTRLFSTD
10  301 WGILGVVWML AVLVLAVNP QRFQDNLVWL LPPLALFGAA QDLSLRGAA
351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAFY SPYYVPDIDP
401 IPMAVAVLET PLWLWATRK NIRGRQAVTN WAAGVTLTWA LLMTLFLFWL
451 DAAKSHAPVVR SMEASLSPE LKRELSGIE CIDIGGGDLH TRIVVTQYGT
501 LPHRVGDVQC RYRIVRLPQN ADAPOGQWTV WQGARPRNKD SKFALIRKGT
15  551 ENILKTTD*

```

ORF141a and ORF141-l show 98.2% identity in 553 aa overlap:

```

or f141a.pep  MLTYTPPDARPPAKTHEKFWLLLLMAFAWLWPGVFSHDLWNPDEPAVYTAVEALAGSPTP
20  or f141-l    MLTYTPPDARPPAKTHEKFWLLLLMAFAWLWPGVFSHDLWNPDEPAVYTAVEALAGSPTP
or f141a.pep  LVAHLFGQIDFGIPPVYLWVAAAFKHLSPWAADPYDAARFAGVFFAVVGLTSCGFAGFN
25  or f141-l    LVAHLFGQIDFGIPPVYLWVAAAFKHLSPWAADPYDAARFAGVFFAVVGLTSCGFAGFN
or f141a.pep  FLGRHHGRSVVLILIGCIGLIPVHFLNPAAAAFAGLVHGYSLARRRVIAASFLLGT
25  or f141-l    FLGRHHGRSVVLILIGCIGLIPVHFLNPAAAAFAGLVHGYSLARRRVIAASFLLGT
or f141a.pep  GWTLMSLAAAPAAAFALMLPLPVLMFFRFWQSRRLMTAVASLAFALPLMTVYPLLAKT
30  or f141-l    GWTLMSLAAAPAAAFALMLPLPVLMFFRFWQSRRLMTAVASLAFALPLMTVYPLLAKT
or f141a.pep  QPALFAQWLDHVGFTGGVVRHIQTAFSLFYYLKNLLWFAFPALPAVMTVCRTRLFSTD
35  or f141-l    QPALFAQWLDYHVGFTGGVVRHVQTAFSLFYYLKNLLWFAFPALPAVMTVCRTRLFSTD
or f141a.pep  WGILGVVWMLAVLVLAVNPQRFQDNLVWLLPPLALFGAAQDLSLRGAAAFVNWFGIMA
40  or f141-l    WGILGVVWMLAVLVLAVNPQRFQDNLVWLLPPLALFGAAQDLSLRGAAAFVNWFGIMA
or f141a.pep  FGLFAVFLWTGFFAMNYGWPAKLAERAAFYSPYYVPDIDPIPMVAVAVLETPLWLWATRK
40  or f141-l    FGLFAVFLWTGFFAMNYGWPAKLAERAAFYSPYYVPDIDPIPMVAVAVLETPLWLWATRK
or f141a.pep  NIRGRQAVTNWAAGVTLTALLMTLFLPWLDAKSHAPVVRSMESLSPELKRELSGIE
45  or f141-l    NIRGRQAVTNWAAGVTLTALLMTLFLPWLDAKSHAPVVRSMESLSPELKRELSGIE
or f141a.pep  CIDIGGGDLHTRIVVTQYGTLPHRVGDVQC RYRIVRLPQNADAPOGQWTVWQGARPRNKD
50  or f141-l    CIDIGGGDLHTRIVVTQYGTLPHRVGDVQC RYRIVRLPQNADAPOGQWTVWQGARPRNKD
or f141a.pep  SKFALIRKIGENI
55  or f141-l    SKFALIRKIGENI

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF141 shows 95% identity over a 140aa overlap with a predicted ORF (ORF141ng) from *N.gonorrhoeae*:

```

60  or f141.pep  DFGISPVYLWVAAAFKHLSPWAADSYDVA 30
or f141ng  WNPAEPAVYTAVEALAGSPTLVAHLFGQIDFGIPPVYLWVAAAFKHLSPWAADHPYDAA 126

```


-341-

orf141.pep	RFAGVFVAVIGLTSCGFAGFNLGRHHGRKVVLILIGICIGLIPVAHFNPAAAAFAAAGL	90
orf141.ng	RFAGVFVAVIGLTSCGFAGFNLGRHHGRSVLILHIGICIGLIPVAHFNPAAAAFAAAGL	186
5 orf141.pep	VLHGYSLARRRVIAASFLTGTWTLMSLAAAYPAAFALMLPLVLMFFFRP	140
orf141.ng	VLHGYSLARRRVIAASFLTGTWTLMSLAAAYPAAFALMLPLVLMFFFRWQSRIMLTA	246

An ORF141ng nucleotide sequence <SEQ ID 599> was predicted to encode a protein having amino acid sequence <SEQ ID 600>:

10	1	MPSEAVSARP	LCEYLLHLAI	RFFLLTLMLT	YTPDPARPPA	KTHEKPWLLL
	51	IMAFANLWPG	VFSDHLWNPA	EPAVYTAVEA	LAGSPTPLVA	HLPGQTFDGI
	101	PPVYLVWAAA	FKELLSPWAA	HPYDAARFAG	VFFAVIGLTS	CGFAGFNFLG
	151	RHHGRSVVLI	HIGICIGLIPV	AHFFNPAAAA	FAAAGLVHNG	YSLARRRVLA
	201	ASFLLTGTWT	LMSLAAAYPA	AFALMLPLV	LMFFRPQSR	RMLTAVASVL
15	251	AFALPLMTVY	PLLLAKTQPA	LEAQWLNHV	PGTFGGVRHI	QRAFSLFHYL
	301	KNLLWFAPPG	LPLAVTVCR	TRLFTDWDGI	LGIVWMLAVL	VLLAENPQRF
	351	QDNLVWLLPP	LALEGAQQLD	SLRRGAAGAV	NWFGIMAEGL	FAVFLWTGFF
	401	AMNYGWPAKL	AEAAAYFSFY	YVPDIDPIM	AVAVLEFPLW	LWATIRKTRN
	451	GRQAVTNWAA	GVTLTWALLM	TLFLPWLDAA	KSHAPVVRSM	EASFSPELKR
20	501	ELSDGIEICG	IGGSDLHTRI	VMTQYGLPH	RVDGRRCRYK	IVRLPQNADA
	551	PQGWQTVWQG	APRNPKDSKF	ALIRKIGENI	LKTTD*	

Further work revealed the following gonococcal DNA sequence <SEQ ID 601>:

25	1	ATGCTGACCT	ATACCCGCCG	CGATGCCCGC	CGCGCCGCCA	AAACCCAGCA
	51	AAAACCGTGG	CTGCTGCTGT	TGATGGCGTT	TGCGTCCGCG	TGGCGCGGCG
	101	TGTTTCCCCA	CGATTGTGGG	AATCCTGCGG	AACCTGCCGT	CTATACCGCC
	151	GTCGAAGCAC	TGGCAGGCAG	CCCCACCCCC	TGGTGTGCCC	ATCTGTTCCG
	201	TCAAAACGAT	TTCCGCATAC	CGCCCGTGTA	TCTTTGGGTT	GCCGCGCGAT
	251	TCAAAACATT	GCTGTCGCCG	TGGGCAGCGC	ACCCGTATGA	TGCCCGCACG
	301	TTTGACAGCG	TATTTTTTGC	CGTTATCGGA	CTGACTCTTT	GCGCGTTTGC
30	351	CGGTTTCAAC	TTTTTGGGCA	GACACACACG	GCGCAGCGTT	GTTTAAATCC
	401	ATATCGGCTG	TATCGGCGTG	ATTCCGGTTG	CCCATTTCTC	CAATCCcgcc
	451	gcgcgcgcct	tTGCcGCcGC	CGGACTGGTG	CTGCACggct	actcgctgGC
	501	ACGCGGCGCG	GATGAtgcgc	cctctTtct	GCTCGGTACG	GSTTGGAGCT
	551	TGATGTCGCT	GCGCGCAGCT	TATCCGCGCG	CGTTTGCCTG	GATGCTGCCC
35	601	CTGCCCGTGC	TGATGTTTTT	CGTCCCGTGG	CAAAACAGCG	GTTTGATGTT
	651	GACGGCAGTC	GCTTCGCTTG	CCTTTGCCCT	CGCGCTTATG	ACCGTTTACC
	701	CGCTGCTCtL	gGCAAAAACG	CAGCCCGCGC	TGTTTGCSCA	ATGGCTCAAC
	751	TATCAAGTTT	TGCGTACGTT	cggcgGCGTG	CGGCAcATTC	AGAGgGCatT
	801	Cagtttctgtt	CactatctgA	AAaactctgt	tlggttcgca	cgcgccgggC
40	851	TGCGCTGGCG	GCTTGGAGCG	GTTTCCCGCG	CACGCGCTGT	CTCGACACAG
	901	TGGGGGATTT	TGGCAGTTGT	CTGGATGCTG	CGCGTTTGGG	TGCTGCTCGC
	951	CTTTAATCCG	CAGCGTTTTC	AAGACAACCT	CGTCTGGCTG	CTCGCGCGCG
	1001	TTGCGCTGTT	CGCGGCGGCG	CAACTGGACA	GCTTGAGGCG	CGGCGCGGCG
45	1051	GCTTTTGTCA	ACTGTTTCGG	CATTATGGCG	TTCGCGCTGT	TTGCGCTGTT
	1101	CGTGTGGACG	GGCTTTTTTG	CCATGAATTA	CGGCTGGCCG	GCCAAGCTTG
	1151	CGAAGCGCGC	CGCCTACTTC	AGCCCGTATT	ACGTTCCCGA	CATCGATCCC
	1201	ATTCCOGATGG	CGGTTGCGGT	ACTGTTCACA	CCCTTGTGGC	TGTGGCGGAT
	1251	TACCCGGAAA	AACATACGGG	CGAGGCGAGG	GTTTACCAAC	TGGGCGCGAG
	1301	CGGTTACCTC	GACCTGGGCT	TGCTGATGAT	CGCTGTTTCT	GCCCTGGCTG
50	1351	GACGCGCGGA	AAAGCCACGC	GCCCCTGCTC	CGGATATATG	AGGCACTGTT
	1401	TTCCCCGGAA	TAAACACGGG	AGCTTTCAAG	CGGCACTGAG	TGTATCGGCA
	1451	TAGCGGCGCG	CGACCTGCAC	ACGCGGATTT	TTTGGACGCA	GTACGCGACA
	1501	TTGCGCGACC	GCGTGGGCGA	TGTCGCTTGC	CGCTACCGTA	TGCTCGCGCT
	1551	GCCCCAAAAC	GCGGATGCGC	CGCAAGGCTG	CGACAGCGCT	TGGCAGGGTG
55	1601	CGCGCCCGCG	CAACAAAGAC	AGTAAGTTTG	CACATGATCG	GAAATACGGG
	1651	GAAATATATAT	TAAAAACAAC	AGATTGA		

This corresponds to the amino acid sequence <SEQ ID 602; ORF141ng-1>:

60	1	MLTYTPPDAR	PPAKTHEKPW	LLLLMAFAML	WPGVFSHDLW	NPAEPVYTA
	51	VEALAGSPTP	LVAILPGQTD	PGIPPPVYLV	AAAFKHLSP	WAADPYDAR
	101	FAGVFFAVIG	LTSCGFAGFN	PLGRHHGRSV	VLILHIGICIG	LIPVAHFNLPA
	151	AAAFPAAGLV	LHGYSLARRR	VIAASFLTGT	WTLMSLAAAY	PAAFALMLPL
	201	LPVLMFFFRW	QSRRLMLTAV	ASLAFALPLM	TVYPLLLAKT	QPALFAQWLN
	251	YHVFGTFGV	RHIQRAFSLF	HYLKNLWFA	PPGLPLAVWT	VCRLTFLSTD
	301	WGILGTVWML	AVLVLLAENP	QRQDNLWLV	LPPLALPGAA	QLDSLRGGA

```

351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAFY SPYYVPDIDP
401 IPMAVAVLFT PLMLWAI TRK NIRGRQAVTN WAAGVTLTWA LLMTLFLPWL
451 DAAKSHAPV RMEASFSPK LKRELSGIE CIGGGGDLH TRIVWTQYGT
501 LPHRVGDVRC RYRVLRLPQN ADAPQGWQTV WQGARPRNKD SKFALIRKIG
551 ENILKTTD*

```

ORF141ng-1 and ORF141-1 show 97.5% identity in 553 aa overlap:

```

      orf141ng-1.pep MLTYTPDPARPPAKTHEKEFWLLMLMAFAWLWPGVFSHDLWNPAEPVYTAVEALAGSPTP
10      orf141-1      MLTYTPDPARPPAKTHEKEFWLLMLMAFAWLWPGVFSHDLWNDEPAVYTAVEALAGSPTP

      orf141ng-1.pep LV AHLFGQTD FGI PPVYLWVAAAFKHLSPWAADPYDAARFAGVFFAVIGL TSCGFAGFN
      orf141-1      LV AHLFGQTD FGI PPVYLWVAAAFKHLSPWAADSYDAARFAGVFFAVIGL TSCGFAGFN

15      orf141ng-1.pep FLGRHHGRSVLIHIGCIGLIPVAHFLNPAFAAAGVLVHGYSLARRRVIAASFLLGT
      orf141-1      FLGRHHGRSVLIHIGCIGLIPVAHFLNPAFAAAGVLVHGYSLARRRVIAASFLLGT

20      orf141ng-1.pep GWTIMSLAAAYPAAFALMLPLPVMFFRQWSRRMLTAVASLAFALPLMTVYPLLAKT
      orf141-1      GWTIMSLAAAYPAAFALMLPLPVMFFRQWSRRMLTAVASLAFALPLMTVYPLLAKT

      orf141ng-1.pep QPALFAQWLDYHVFGT FGGVRHVIQAFSLFHYLKNLLWFALPALPLAVMTVCRTLSTFD
25      orf141-1      QPALFAQWLDYHVFGT FGGVRHVIQAFSLFHYLKNLLWFALPALPLAVMTVCRTLSTFD

      orf141ng-1.pep WGILGIVWMLAVLVLLAFNPQRFDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA
      orf141-1      WGILGIVWMLAVLVLLAFNPQRFDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA

30      orf141ng-1.pep FGLFAVFLWTGFFAMNYGWPAKLAERAAFYSPYYVPDIDP IPMAVAVLFTPLMLWAI TRK
      orf141-1      FGLFAVFLWTGFFAMNYGWPAKLAERAAFYSPYYVPDIDP IPMAVAVLFTPLMLWAI TRK

35      orf141ng-1.pep NIRGRQAVTNWAAGVTLT WALLMTLFLPWLDAKSHAPVVRSMESFSPKRELSGIE
      orf141-1      NIRGRQAVTNWAAGVTLT WALLMTLFLPWLDAKSHAPVVRSMESFSPKRELSGIE

40      orf141ng-1.pep CIGGGGDLHTRIVWTQYGT LPHRVGDVRCRYRIVRLPQNADAPQGWQTVWQGARPRNKD
      orf141-1      CIGGGGDLHTRIVWTQYGT LPHRVGDVRCRYRIVRLPQNADAPQGWQTVWQGARPRNKD

      orf141ng-1.pep SKFALIRKIGENILKTTDX
45      orf141-1      SKFALIRKIGENIX

```

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 72

50 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 603>:

```

1  ..CAATCCGCCA AATGGTTATC GGGCCAAACT CTAGTCGGCA CAGCAATTGG
51 GATACGCGGG CAGATAAGC TTGGCGGCCA CCTGCATTAC GATATATTTA
101 CCGGCCGCGC ATTGAAAGG CCGCAATTTT TCCAATCAAG GAATGGGCA
151 AGCGGTTTT CAGTAGGCTA TACGTTTTTA

```

55 This corresponds to the amino acid sequence <SEQ ID 604; ORF142>:

```

1  ..QSAAWLSGQT LVGTAGIRG QIKLGNLHY DIPTGRALK PEFFQSRKWA
51 SGFQVGYTF*

```

Further work revealed the complete nucleotide sequence <SEQ ID 605>:

```

1  ATGGATAATT  CGGCTAGTGA  GCGGACAGGA  AAATACCAAG  GAAATATCAC
51  TTTCTCTGCC  GACAATCCTT  TGGAGCTGAG  TGATATGTTT  TATGTAAATT
101  ATGGACGTTT  GATTGGCGGT  ACGGCCGATG  AGGAAAGTTT  TGACGGCCAT
151  CGCAAGAAG  GCGGATCAAA  CAATTACGCC  GTACATTATT  CAGCCCTTTT
201  CGGTAAATGG  ACATGGGCTT  TCAATCACAA  TGGCTACCGT  TACCATCAGG
251  CAGTTTCCGG  ATTATCGGAA  GTCTATGACT  ATAATGAAAA  AAGTTACAAT
301  ACTGATTTCG  GCTTCAACCG  CTTGTGTAT  CTTGATGCCA  AACGCAAAAC
351  CTATCTCGGT  GTAAAACCTG  GGATGAGGGA  AACAAAAGT  TACATTGATG
401  ATGCGGAAC  GACTGTACAA  CGCGTAAAA  CTCGGGGTTG  GTTGGCAGAA
451  CTTTCCACA  AAGAAATAT  CGGTGCGAGT  ACGGCAGATT  TTAAGTTGAA
501  ATATAAACGC  GGCACCGGCA  TGAAGAATGC  TCTCGCGCG  CCGTGAAGAAG
551  CCTTTGGGGA  AGGCACGTCA  CGTATGAAAA  TTGGACGCG  ATCGGCTGAT
601  GTAAATATCT  CTTTTCAAAT  CGGTAAACAG  CTATTGCTCT  ATGACACATC
651  CGTTCATGCA  CAATGGAACA  AAACCCGCT  AACATCGCAA  GACAACTGG
701  CTATCGCGGG  ACACACACAC  TACGTGGCT  TCGACGGTGA  AATGAGTTTG
751  TCTGCCGAG  GGGGATGTTA  TTGGCGCAAC  GATTGAGCT  GGCAATTTAA
801  ACCAGCCAT  CAGCTTATCT  TTGGGCTGA  TTGAGGACAT  GTTTCAGGAC
851  AATCGGCCAA  ATGTTTATCG  GGCACAACT  TACTCGGCAC  AGCAATTGGG
901  ATACGCGGG  AGATAAAGCT  TGGCGGCAAC  TCGCATTAG  ATATATTAC
951  CGGCGCGGCA  TTGAAAAGC  CGGAATTTT  CCAATCAAG  AAATGGGCAA
1001  GCGGTTTTCA  GGTAGGCTAT  ACGTTTTAA

```

This corresponds to the amino acid sequence <SEQ ID 606; ORF142-1>:

```

1  MDNSGSEATG  KYQGNITFSA  DNPLGLDMF  YVNYGRSIGG  TPDEESFDGH
51  RKEGGSNNYA  VHSYAPFFKW  TWAFNHNGYR  YHQVAVSLSE  VVDYNGKSYN
101  TDFGNRLLY  RDAKRKTYLG  VKLWMRETKS  YIDDAELTVQ  RKRATGWLAE
151  LSHKEYIGRS  TADFKLYK  GTMKDALRA  PEEAFGGTGS  RMKIWTASAD
201  VNTFPQIGKQ  LFAYDTSVHA  QWNKPTLSQ  DKLAIGGHT  VRGFDGEMSL
251  SAERGWYWRN  DLWQFKFGH  QLYLGADVGH  VSGQSAWLKS  GQTLVGTAGI
301  IRGQIKLGGN  LHYDIFTGRA  LKPEFFQSR  KWASGFQVGY  TE*

```

30 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF142 shows 88.1% identity over a 59aa overlap with a predicted ORF (ORF142ng) from *N.gonorrhoeae*:

```

35  orf142.pep  QSAKWLSGQTLVGTAGIRGQIKLGGNLYH  30
      orf142ng  RGWYWRNDLSWQFKPGHQLVLGADVGHVSGSAKWLSGQTLGATGIRGQIKLGGNLEY  313
      orf142.pep  DIFTGRALKKPEFFQSRKWASGFQVGYTF  59
      orf142ng  DIFTGRALKKPEYFQTKKWTGFGVGYSF  342

```

The complete length ORF142ng nucleotide sequence <SEQ ID 607> is:

```

1  ATGGATAATT  CGGCTAGTGA  GCGGACAGGA  AAATACCAAG  GAAATATCAC
51  TTTCTCTGCC  GACAATCCTT  TGGAGCTGAG  TGATATGTTT  TATGTAAATT
101  ATGGACGTTT  GATTGGCGGT  ACGGCCGATG  AGGAAAGTTT  TGACGGCCAT
151  CGCAAGAAG  GCGGATCAAA  CAATTACGCC  GTACATTATT  CAGCCCTTTT
201  CGGTAAATGG  ACATGGGCTT  TCAATCACAA  TGGCTACCGT  TACCATCAGG
251  CCGTTTCCGG  ATTATCGGAA  GTCTATGACT  ATAATGAAAA  AAGTTACAAT
301  ACTGATTTCG  GCTTCAACCG  CTTGTGTAT  CTTGATGCCA  AACGCAAAAC
351  CTATCTCGGT  GTAAAACCTG  GGATGAGGGA  AACAAAAGT  TACATTGATG
401  ATGCGGAAC  GACTGTACAA  CGCGTAAAA  CCACAGGTTG  GTTGGCAGAA
451  CTTTCCACA  AAGAAATAT  CGGTGCGAGT  ACGGCAGATT  TTAAGTTGAA
501  ATATAAACAC  GGCACCGGCA  TGAAGAATGC  TCTCGCGCG  CCGTGAAGAAG
551  CCTTTGGGGA  AGGCACGTCA  CGTATGAAAA  TTGGACGCG  ATCGGCTGAT
601  GTAAATATCT  CTTTTCAAAT  CGGTAAACAG  CTATTGCTCT  ATGACACATC
651  CGTTCATGCA  CAATGGAACA  AAACCCGCT  AACATCGCAA  GACAACTGG
701  CTATCGCGGG  ACACACACAC  TACGTGGCT  TCGACGGTGA  AATGAGTTTG
751  CCTGCCGAG  GGGGATGTTA  TTGGCGCAAC  GATTGAGCT  GGCAATTTAA
801  ACCAGCCAT  CAGCTTATCT  TTGGGCTGA  TTGAGGACAT  GTTTCAGGAC
851  AATCGGCCAA  ATGTTTATCG  GGCACAACT  TACTCGGCAC  AGCAATTGGG
901  ATACGCGGG  AGATAAAGCT  TGGCGGCAAC  TCGCATTAG  ATATATTAC
951  CGGCGGTGCA  TTGAAAAGC  CGGAATTTT  TCAGACGAAG  AATGGGTTAA

```

1001 CGGGGTTTCA GGTGGGTTAT TCGTTTTGA

This encodes a protein having amino acid sequence <SEQ ID 608>:

1 MDNSGSEATG KYQGNITFSA DNPFLSDMF VYNYGRSIGG TPDEENFDGH
5 51 RKEGGSNNYA VHSAPFGKW TWFNHNHGYR YHQA VSGLSG VYDYNKSYN
101 TDFGENRLLY RDAKRRTYLS VKLWTRTKS YIDDAELTVQ RRKTGWLAEL
151 LSHKGYIGRS TADFKLYKHK GTGMKDALRA PEEAFEGETS RMKIWTASAD
201 VNTFPQIGKQ LFAYDTSVHA QWNKTPLTSQ DKLAIGGHHT VRGFDGEMSL
251 PAERGWYWRN DLSWQFKPGH QLYLGADVGH VSGQSAKWL GQTLAGTAIG
301 IRGQIKLGGN LHYDIFTGRA LKKPEYFQTK KWTGFGVGY SF*

10 The underlined sequence (aromatic-Xaa-aromatic amino acid motif) is usually found at the C-terminal end of outer membrane proteins.

ORF142ng and ORF142-1 show 95.6% identity over 342aa overlap:

orfl42-1.pep MDNSGSEATGKYQGNITFSADNPLGLSDMFVYNYGRSIGGTPDEESFDGHRKEGGSNNYA
15 orfl42ng-1 MDNSGSEATGKYQGNITFSADNPLGLSDMFVYNYGRSIGGTPDEENFDGHRKEGGSNNYA
orfl42-1.pep VHSAPFGKWTFWFNHNHGYRYHQA VSGLSEVYDYNKSYNTDFGNRLLYRDAKRRTYLS
20 orfl42ng-1 VHSAPFGKWTFWFNHNHGYRYHQA VSGLSEVYDYNKSYNTDFGNRLLYRDAKRRTYLS
orfl42-1.pep VKLWMTREKSYIDDAELTVQRRKTAGWLAELSHKEYIGRSTADFKLYKRGTMKDALRA
orfl42ng-1 VKLWMTREKSYIDDAELTVQRRKTAGWLAELSHKEYIGRSTADFKLYKRGTMKDALRA
25 orfl42-1.pep PEEAFEGETS RMKIWTASADVNTFPQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHT
orfl42ng-1 PEEAFEGETS RMKIWTASADVNTFPQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHT
orfl42-1.pep VRGFDGEMSLAERGWYWRN DLSWQFKPGHQLYLGADVGHVSGQSAKWLGGQTLVGTAG
30 orfl42ng-1 VRGFDGEMSLAERGWYWRN DLSWQFKPGHQLYLGADVGHVSGQSAKWLGGQTLVGTAG
orfl42-1.pep IRGQIKLGGNLYHYDIFTGRALKKPEFFQSRKWASGFGVGYTF
35 orfl42ng-1 IRGQIKLGGNLYHYDIFTGRALKKPEYFQTKKWTGFGVGYSF

In addition, ORF142ng is homologous to the HecB protein of *E. chrysanthemi*:

gi|1772622 (L39897) HecB [Erwinia chrysanthemi] Length = 558
Score = 119 bits (295), Expect = 3e-26
Identities = 88/346 (25%), Positives = 151/346 (43%), Gaps = 22/346 (6%)
40 Query: 2 DNSGSEATGKYQGNITFSADNPLGLSDMFVYNYGRSIGGTPDEENFDGHRKEGGSNNYAV 61
Score: 119 bits (295), Expect = 3e-26
Identities = 88/346 (25%), Positives = 151/346 (43%), Gaps = 22/346 (6%)
Sbjct: 230 DNSGSEATGKYQGNITFSADNPLGLSDMFVYNYGRSIGGTPDEENFDGHRKEGGSNNYAV 61
45 Query: 62 HYSAPFGKWTFWFNHNHGYRYHQA VSGLSEVYDYNKSYNTDFGNRLLYRDAKRRTYLSV 121
+S P+G W +N++ RY + G S F +R+++RD RT ++
Sbjct: 281 +FMPYGYWNLGYNYSQSRYRNTINRDPWHTGSDSTHRFSLSRVGVDTGTMKTAIG 339
Query: 122 KLWMTREKSYIDDAELTVQRRKTAGWLAELSHKEYIGRSTADFKLYKRGTMKDALRAP 181
R +Y++ +L RK + ++H + A F Y G +
50 Sbjct: 340 TFSQRTGNNYLLNGSLPSSSRKLLSSVSLGVNHSQKLMGGLATFNPTNYRGRVWLGSETDT 399
Query: 182 EEAFFEGETS RMKIWTASADVNTFPQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHTV 241
+++ E + WT SA P Y S++ Q++ L ++L +GG ++
55 Sbjct: 400 DKSADPEPRAEFNKWTLASASYHPV---TDSITLYGLSGQY SARALYVSGEQLTLGGESSI 456
Query: 242 RGF DGMESLPAERGWYWRN DLSWQFKPGHQLYLGADVGHVSGQSAKWLGGQTLVGTAG 296
RGF E RG YWRN++WQ G+ ++ A D GH+ + +L G
Sbjct: 457 RGF-REQYTSGNRGAYWRNENLWQAWQLPVLGNVTFMAADVGGHLYNHKQDNSTAA SLWG 515
60 Query: 297 TAIGIRGQIKLGGNLYHYDIFTGRALKKPEYFQTKKWTGFGVGYSF 342
AtG+ + L + G + P + Q V G++V G SF


```

orf143a      YGDENGSDGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLANANFHHEAAEELGLLAAE
            80      90      100      110      120      130
5      orf143.pep      VAQMEKKYRLLIKNN
            ||||| ||||| |||||
orf143a      VAQMEKKYRLXIKNNLYINNNAGVCDPSGQSELTFFPLYIGSTKFILVIGGIPDLGKEA
            140      150      160      170      180      190

```

The complete length ORF143a nucleotide sequence <SEQ ID 613> is:

```

10      1      ATGGAATCAA  CANTTTCAC  ACAAGCAAA  TTATATCNCC  GCCTGACTCC
51      TGCCGGTGCA  TTTTATGCCG  TATCCAGCGA  TGNCCCCAGT  GCCGTAATAA
101     CTTTGTGTGCA  CAGCCTGTGT  AAGCGGATG  CGGACGAAAT  GGTNAGCAGT
151     GAGRAGCTGC  TTACCTGGGC  GGANACCGCC  GACATCGATA  CCGCTTTGAA
201     CCTGTTGTAC  CGTTTGCAA  AACTCGAATT  CCTCTATGGC  GATGAAAACG
15      251     GTCATTGAGA  CGGCATCAAT  TTGTCGAGC  AGCAATTGCC  GTTGTCTGAT
301     GAACAATTGT  CCGGCAGCGG  TAAGCGGTTA  TTGGTCGATC  GGAACGGTCT
351     GTATCTTGCC  AACGCCAATT  TCCATCATGA  GCGCGCGGAA  GAGTTGGGGT
401     TGTGCGCGGC  AGAAGTCGCA  CAGATGGA  AGAAATACCG  GCTGNNATT
20      451     AAGAACACC  TGTATATCAA  CATTACGCT  TGGGCGCTTT  CGCATCTCT
501     CGGTACAGGC  GAATGACAT  TTTCACCAT  GTATATCGGT  TCACCAAAAT
551     TTTATTTGGT  TATCGCGCGC  ATTCCGATT  TGGGCAAGA  GGCATTGTGT
601     ACTTTGGTAA  GGATNTTATA  CMCNCNGTA  CAGCAACCGC  GTGTAAAAC
651     TGGGAGAGAG  GANGGTTAT  GCAGCAATTA  TTGA

```

This encodes a protein having amino acid sequence <SEQ ID 614>:

```

25      1      MESTXSLQAN  LYXRLTPAGA  FYAVSSDXPS  AGKTLHLSLL  KADADEMVSS
51      EKLLTWAXTA  DIDTALNLLY  RLQKLEFLYG  DENGSDGIN  LSDEQLPLLM
101     EQLSGSGKAL  LVDNRNGLYLA  NNFHHEAAE  ELGLLAAEVA  QMEKKYRLXI
151     KNNLYINNA  NGVCDPSGQS  ELTFFPLYIG  STKFILVIGG  IPDLGKEAFV
201     TLVRXLYXLL  QPRVKLGRE  XGLCSNY*

```

30 ORF143a and ORF143-1 show 97.1% identity in 207 aa overlap:

```

orf143a.pep      MESTXSLQANLYXRLTPAGAFYAVSSDXPSAGKTLHLSLLKADADEMVSSSEKLLTWAXTA
orf143-1          MESTLSLQANLYPRLTTPAGAFYAVSSDAPSAKTLHLSLLKADADEMVSSSEKLLTWADTA
35      orf143a.pep      DIDTALNLLYRLQKLEFLYGDENGSDGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA
orf143-1          DIDTALNLLYRLQKLEFLYGDENGSDGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA
40      orf143a.pep      NNFHHEAAEELGLLAAEVAQMEKKYRLIKNNLYINNNAGVCDPSGQSELTFFPLYIG
orf143-1          NNFHHEAAEELGLLAAEVAQMEKKYRLIKNNLYINNNAGVCDPSGQSELTFFPLYIG
orf143a.pep      STKFILVIGGIPDLGKEAFVTLVRXLY
orf143-1          STKFILVIGGIPDLGKEAFVTLVRILY
45

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF143 shows 95.5% identity over a 110aa overlap with a predicted ORF (ORF143ng) from *N.gonorrhoeae*:

```

50      orf143.pep      MRTKWSAVRSCTWADTADIDTALNLLYRLQKLEFLYGDENGSDGINLSDEQLPLLMEQL      60
orf143ng          MRTKWSAVRSCSRADTADIDTALNLLYRLQKLEFLYGDENGSDGINLSDEQLPLLMEQL      60
orf143.pep      SSGSKALLVDRNGLYLANANFHHEAAEELGLLAAEVAQMEKKYRLLIKNN      110
orf143ng          SSGSKALLVDRNGLYLANANFHHEAAEELGLLAAEVAQMEKKYRLLIKNNLYINNNAGV      120

```

An ORF143ng nucleotide sequence <SEQ ID 615> was predicted to encode a protein having amino acid sequence <SEQ ID 616>:

```

1 MRTKWSAVRS CSRADTADID TALNLLYRIQ KLEFLYGDEN GHSDDGINLSD
51 EQLPLMEQL SSGSKALLVD RNLGYLANAN FHESAEELG LLAAEVAQME
101 KKYRLIRNN LYINNNANGV CDPSPGQSELT FFPLYIGSTK FILVIAGIPD
151 LSKGGICYFG KDFIPPLQVP RVKLGTTGIM RLLISLLED LNNSTDTIA
5 SAVISTDGLP MATMLPSHLN SDRVGAISAT LALGSRSQV ELACGELEGV
251 MIKKGSGYIL LSGAQDAVL VLVAKETGRL GLILLDAKA ARHIAEAI*

```

Further work revealed the following gonococcal DNA sequence <SEQ ID 617>:

```

1 ATGGAATCAA CACTTTCACT ACAAGCGAAT TTATATCCCT GCCTGACTCC
51 TGCCGGTGCA TTTTATGCGG TATCCAGCGA TGCCGCCAGT GCCGGTAAAA
101 CTTTGTTCGG CAGCGCTGTT AAGCGGATG CGGACGAAGT GSTCAGCAGT
151 GAGAAGCTGC TCGCGGCGGA CACCGCGGAC ATCGATACCG CTTTGAACCT
201 GTTGTACCGT TTGCAAAAAC TCGAATTCCT CTATGGCGAT GAAAACGGTC
251 ATTCAGACGG CATCAATTG TCGGACGAGC AATTGCCGTT GCTGATGGAA
301 CAATTGTCGG GCAGCGGTAA GGCATTATTG GTCGATCGGA ACGGTCTGTA
151 TCTTGCCAAC GCCAATTTC ATCATGAGTC GCGGGAAGAG TTGGGGTTGT
401 TGCGGCGAGA AGTCGCACAG ATGGAAGAAG AATACCGGCT GCTGATTAGG
451 AACAGCTGT ATATCAACAA TAACGCTTGG GCGGTTTGGC ATCCTTCCGG
501 TCAGAGCGAA TTGACATTTT TCCCATTTGA TATCGGTTCA ACCAAATTTA
551 TTTTGGTTAT CGCGCGCAT CCGGATTTG GCAAGAGAGC ATTTGTACT
201 TTGTAAGGA TTTTATACCG CCGTTACAGC AACCGCGTGT AA

```

This corresponds to the amino acid sequence <SEQ ID 618; ORF143ng-1>:

```

1 MESTLSLOAN LYPCLTPAGA FYAVSSDAPS AGKTLRLSLL KADADEVVSS
51 EKLLAADTAD IDTALNLLYR LQKLEFLYGD ENGHSDGINL SDEQPLIME
101 QLSGSGKALL VDRNGLYLAN ANFHESAEEL LGLLAAEVAQ MEKKYRLLLR
151 NNLYINNNAW GVCDPSGQSE LTFFPLYIGS TKFILVIAGI PDLSCKEAFVT
201 LVRILYRRYS NRV*

```

ORF143ng-1 and ORF143-1 show 95.8% identity in 214 aa overlap:

```

orf143ng-1.pep MESTLSLOANLYPCLTPAGAFYAVSSDAPSAGKTLRLSLLKADADEVVSSSEKLLA-ADTA 59
30 orf143-1 MESTLSLOANLYPRLTPAGAFYAVSSDAPSAGKTLRLSLLKADADEVVSSSEKLLTWADTA 60
orf143ng-1.pep DIDTALNLLYRQKLEFLYGDENGHSDGINLSDEQPLIMEQLSGSGKALLVDRNGLYLA 119
35 orf143-1 DIDTALNLLYRQKLEFLYGDENGHSDGINLSDEQPLIMEQLSGSGKALLVDRNGLYLA 120
orf143ng-1.pep NANFHESAEELGLLAAEVAQMEKKYRLLIRNNLYINNNANGVCDPSGQSELTFFPLYIG 179
orf143-1 NANFHESAEELGLLAAEVAQMEKKYRLLIRNNLYINNNANGVCDPSGQSELTFFPLYIG 180
40 orf143ng-1.pep STKFILVIAGIPDLSCKEAFVTLVRILYRRYSNRV 213
orf143-1 STKFILVIAGIPDLSCKEAFVTLVRILYRRYSNRV 214

```

Based on the presence of the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 74

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 619>:

```

1 ATGACCTTTT TACAAAGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
51 GTTTGCTATGG TTGCTGCTCC GCGCGTTTGA TGAAGAACGC GTACCGCAGT
101 CGCGGCGCAAG CATGAAGTTT AGACGCGTGG TGCGACTCTGT CCGCGTGTCTG
151 ACGGTGATGG TGCGCGTGGC TTCGATTTTC CCGGTGTTGG ACGCGTGGTC
201 GGATTCGTTT GTCTCTCTGG TCAACCAAAAC CATTTGSGCG CA_GCGCGGG
251 ACATGCTGTT CGACTATATG AATGCGTTTC GCGAGCAGGC GAACCGCGCT
301 ACGCATCTGC GCAGCGTGAT GCTGGTCTGT ACCTGCGTGA TGTGATTTG
351 GACGATAGAC AATACGTTC ACGCATCTGC GCGGCGTCAA WTYCGCGCGT
401 CGTGGATGAT..

```

This corresponds to the amino acid sequence <SEQ ID 620; ORF144>:

```

1  MTFLQRLQGL ADNKAFAW FVVRFFDEER VPQAAASMTF TLLALVPVL
51  TVMVAVASIF PVFDRWSDSF VSFVNQTFV XGADMVFDYI NAFREQANRL
101 TAIGSVMLVV TSLMLRTID NTFNRIWRVX XQRPWM...

```

5 Further work revealed the complete nucleotide sequence <SEQ ID 621>:

```

1  ATGACCTTTT TACAAGCTTT GCAAGGTTTG CGAGACAATA AATCTGTGC
51  GTTTGCATGG TTGCTGCTCC GCGCGTTTGA TGAAGAACGC GTACCGCAGG
101 CGCGGGCAAG CATGACGTTT ACGACGCTGC TGGCACTCGT CCGCGTCTGT
151 ACCGCTGATG TGGCGGTGCG TTGCAATTTG CCGGTGTTCC ACCGCTGGTC
201 GGATTGCTTC GTCTCTCTCG TCAACCAAC CATTTGTCGC CAGCGCGCG
11  251 ACATGGTGTG CGACTATATC AATGCGTTCC GCGAGGAGCG GAACCGCTG
301 ACGGCAATCG CGACGCTGAT GCTGTGCTGT ACCTCGCTGA TCGTGATTC
351 GACGATAGAC AATACGTTCA ACCGCATCTG CCGGGTCAAT TCCAGCGCT
401 CGTGGATGAT GCGATTTCTC GTCTATTGGG CTTTACTGAC GTTCGGGCGG
12  451 CTGCTTTTGG CGGTGGGCAT TTCTTTATG GTCCGCTCGG TACAGGATCG
501 CGCGCTTGCC TCAGGTGCGC CGCAGTGTGT GGGCGGCTTG CGAACCGCGG
551 CGACGCTGAC CTTCTAGACG CTTTGTCTGT GGGGGCTGTA CGCTCTCGT
601 CCAACCGCTC TCGTCCCCTG GCGGCAGCGG TTTGTGCGGG CTTTGGCAAC
651 AGCGTTTTTG CTGGAACCGC CGCGCTCCCT CTTCACTTGG TATATGGGCA
701 ATTTGACGCG CTACCGCTCG ATTTACGGCG CGTTTGGCCG CGTGCCTTTT
751 TTTCTGTGTG GCGTGAACCT GTTGTGACG CTGCTCTTGG GCGCGCGGCT
801 GCTGACTTCT TCACTCTCCT ACTGCGAGG AGAAGCGTTC CGCAGGGGCT
851 TCGACTCGCG CGGACGCTTT GACGACGCTG TGAAATTCCT GCTGCTTCTG
901 GATGCGGCGC AAAAAGAGGG CAAAGCCTTG CTTGTTGAGG AGTTCAGAGC
951 GCATATCAAT ATGGGCTACG ACAGTGTGG CGAGCTTTTG GAAGAAGTGG
101 1001 CGCGGCACGG CTACATCTAT TCCGCGACG AGGCTTGGGT GTTGAACAAC
1051 GGGCGGGAAT CGATTGAGTT GAAACGACT TCAAGCTCT TCGTTACCG
1101 TCCGTTGCCT GTGGAAGGG ATCATGTGAA CCAAGCTGTC GATGGGTAA
1151 TGACACGCTG TTTGACAGCT TTGAACATGA CGCTGGCAGA GTTTGACGCT
1201 CAGGCGAAAA AACGGCAGTA G

```

This corresponds to the amino acid sequence <SEQ ID 622; ORF144-1>:

```

1  MTFLQRLQGL ADNKAFAW FVVRFFDEER VPQAAASMTF TLLALVPVL
51  TVMVAVASIF PVFDRWSDSF VSFVNQTFV XGADMVFDYI NAFREQANRL
101 TAIGSVMLVV TSLMLRTID NTFNRIWRVX SQRPMMQFL VYKALLTFGP
151 LSLGVGISFM VGSVQDAALA SGAPQWSGAL RTAATLFTTF LLLWGLYRFV
201 PNRFPARQA FVGALATAFC LETARSLFTW YMGNFDGYRS IYGAFAVPF
251 FLLNLLNLT LVLGGVLTLS SLSYWGGEAF RGFDSRGFR DDVLLILL
301 DAAQKEGKAL PVQEFRRHIN MGYDELGELL EKLAHGYIY SGROGWVLT
351 GADSIKLNEL FKLFVYRFLP VERDHVNAQV DAVMTPLQIT LNMTLAEFDA
401 QAKRK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF144 shows 96.3% identity over a 136aa overlap with an ORF (ORF144a) from strain A of *N.*

meningitidis:

```

45  orf144.pep      10      20      30      40      50      60
      MTFLQRLQGLADNKAFAWFVVRFFDEERVPOXAAASMTFTTLLALVPVLTVMVAVASIF
      |||
orf144a      MTFLQRLQGLADNKAFAWFVVRFFDEERVPOXAAASMTFTTLLALVPVLTVMVAVASIF
      |||
      10      20      30      40      50      60
50  orf144.pep      70      80      90      100     110     120
      PVFDRWSDSFVSFVNQITVPXGADMVFDYINAFREQANRLTAIGSVMLVVTSIMLIRTD
      |||
orf144a      PVFDRWSDSFVSFVNQITVPXGADMVFDYINAFREQANRLTAIGSVMLVTSXMLIRTD
      |||
      70      80      90      100     110     120
55  orf144.pep
      130
orf144.pep      NTFNRIWRVXXQRPWM
      |||
60  orf144a      NTFNRIWRVNSQRPMQFLVYKALLTFGLSLGVGISFXVGSVQDAALASGAPQWSGAL

```


130 140 150 160 170 180

The complete length ORF144a nucleotide sequence <SEQ ID 623> is:

1 ATGACCTTTT TACAAGCTTT GCAAGGTTTG GCAGACAATA AATCTGTGC
 51 GTTTCATAGG TTGCTGTGCC GCGCTTTGA TGAAGAACC GTACCGCAGG
 101 CGGCGCAAG CATGACGTTT ACGACACTGC TGGCCTCTGT CCCCCTGCTG
 151 ACCGTGATGG TGGCGGTGCG TTGATTTTC CCGCTGTGTG ACCGNTGGTC
 201 GGATTCGTTT GTCTCTCTCG TCAACCAAAAC CATTTGTGCC GAGGCGCGCG
 251 ACATGGTNTT CGACTATATC AATGCGTTCC GCGAGCAGCG GAACCGGCTG
 301 ACGGCAATCG GCAGCGTGAT GCTGCTGCTT ACCTCGCNGA TGCTGATTGC
 351 GACGATAGAC AATACGTTCA ACGCATCTGC GCGGCTCAAT TCCACGCGTC
 401 CGTGGATGAT GCAGTTTCTC GTCTATTGGG CTTTACTGAC GTTCGGGCGG
 451 CTGCTTTTGG GCGTGGGCAT TTCCTTTATN GTCGGCTCGG TACAGGATGC
 501 GCGGCTTTGCC TCAGGTGGCG CGCAGTGGTC GGGCGGCTTG CGAAOCCGGC
 551 CGACGCTGAN CTTCATGACG CTTTGTCTGT GGGGGCTGTA CGCTNCGTGT
 601 CCAAAACGCT TCGTTCGCCG GCGGCGANGC TTTGTCCGGG CTTTGCACAC
 651 AGCGTCTGTG CTGGAACCG GCGCTTCCCT CTTTACTTGG TATATGGGCA
 701 ATTTGACGCG CTACCGCTCG ATTACGGNG CGTTTGCCTG CGTGCGGTTT
 751 TTTCTGTTGT GCGCTAACCT GTTGTGAGCG CTGGCTTGG GCGGCGCGGT
 801 GCTGACTTCT TCCTCTCTCT ACTGCGAGG AGAAGCGTTC CGCAGGNGCT
 851 TCGACTCGCG CGAGCGGTTT GAGCAGCTGT TGAATATCT GCTGCTCTG
 901 GATGCGCGCG AAAAAGGAG CNAAGCCTTG CTTGTCCAG AGTTCAGACG
 951 GCATATCAAT ATGCGCTACG ACGAGTTGGG CGAGCTTTTG GAAAGACGCG
 1001 CGCGGCGACG CTACATCTAT TCCGCGACAG AGGTTGGGT GTTGAAGACG
 1051 GGGCGCGATT CGATTGAGTT GAACGAACCT TTCAAGCTCT TCGTTTACCG
 1101 TCGGTTGCCT GTGGAAGGAG ATCATGTGAA CCAAGCTGTG GATGCGGTAA
 1151 TGATGCGGTG TTTGCAGACT TTGACATGA CGCTGCGAGA GTTTGACGCT
 1201 CAGGCGAAAA AACAGCAGCA ATCTTGA

This encodes a protein having amino acid sequence <SEQ ID 624>:

1 MTFQLRQLGL ADNKAICAFW FVVRFRDEER VPQAAASMT TLLALVPVL
 51 TVMVAVASIF PVFDRWSDSF VSEFVNQITV QGADMVFDYI NAFREQANRL
 101 TAIGSVMLVV TSMXLRTID NTFNRIWRVN SQRFMMQFL VYVALLTFPG
 151 LSLGVGISFX VGSVQDAALA SGAPQWSGAL RTAATLXFMT LLLMGLYRXV
 201 PNRFPVARKA FVGALATAFC LETARSLFTW YMGNFDDYRS IYGAFAAVPF
 251 FLLMLNLWT LVLGGAVLTS SLSYWQGEAF RRFDSRGRF DDVLLKILLL
 301 DAAQKEGKAL PVQFRHIN MGVDLGEILL EKLAIRHGYIY SGRQGWVLT
 351 GADSIENEL FKLFVYRPLP VERDHNQAV DAVMPCLOT LNMTLAEFDA
 401 QAKKQQQS*

ORF144a and ORF144-1 show 97.8% identity in 406 aa overlap:

40 orf144a.pep MTFQLRQLGLADNKAICAFWFVVRFRDEERVPQAAASMTFTLLALVPVLTVMVASIF
 orf144-1 MTFQLRQLGLADNKAICAFWFVVRFRDEERVPQAAASMTFTLLALVPVLTVMVASIF
 45 orf144a.pep PVFDRWSDSFVSEFVNQITVPGADMVFDYINAFREQANRLTAIGSVMLVVTSMXLRTID
 orf144-1 PVFDRWSDSFVSEFVNQITVPGADMVFDYINAFREQANRLTAIGSVMLVVTSMXLRTID
 50 orf144a.pep NTFNRIWRVNSQRFMMQFLVYVALLTFGPLSLGVGISFXVGSVQDAALASGAPQWSGAL
 orf144-1 NTFNRIWRVNSQRFMMQFLVYVALLTFGPLSLGVGISFMVGSVQDAALASGAPQWSGAL
 55 orf144a.pep RTAATLXFMTLLMGLYRXVFNRFVPAKXAFVAGALATFCELETARSLFTWYMGNFDDYRS
 orf144-1 RTAATLXFMTLLMGLYRFVFNRFVPAKXAFVAGALATFCELETARSLFTWYMGNFDDYRS
 60 orf144a.pep IYGAFAAVPFLLMLNLWTVLGGAVLTSLSYSWQGEAFRRXFDSSRGRFDDVLLKILLL
 orf144-1 IYGAFAAVPFLLMLNLWTVLGGAVLTSLSYSWQGEAFRRGFDSRGRFDDVLLKILLL
 65 orf144a.pep DAAQKEGKALPVQFRHINMGYDELGEILLEKLAIRHGYIYSGRQGWVLTGADSIENEL
 orf144-1 DAAQKEGKALPVQFRHINMGYDELGEILLEKLAIRHGYIYSGRQGWVLTGADSIENEL
 orf144a.pep FKLFVYRPLPVERDHNQAVDAVMPCLOTLNMTLAEFDAQAKKQQQS 408
 orf144-1 FKLFVYRPLPVERDHNQAVDAVMPCLOTLNMTLAEFDAQAKKQQ 406

Homology with a predicted ORF from *N.gonorrhoeae*

ORF144 shows 91.2% identity over a 136aa overlap with a predicted ORF (ORF144ng) from *N.gonorrhoeae*:

5	orf144.pep	MTFLQRLQGLADNKICAFAMFVVRVFEDEVPQXAASMTFTTLLALVPVLTVMVAVASIF	60
	orf144ng	MTFLQCMQGSADNKICAFAMFVIRRFSEERVPQXAASMTFTTLLALVPVLTVMVAVASIF	60
10	orf144.pep	PVFDRWSDSEFVSFNQITVPXGADMVFDYINAFREANRLTAIGSMVLVVTSLMLIRTID	120
	orf144ng	PVFDRWSDSEFVSFNQITVPXGADMVFDYIDAFRDQANRLTAIGSMVLVVTSLMLIRTID	120
	orf144.pep	NTFNRIWRVXQRPWM	136
15	orf144ng	NAFNRIWRVNTQRFWMQFLVYVWALLTFGPISLGVGISFMVGSVQDSVLSSGAQWADAL	180

The complete length ORF144ng nucleotide sequence <SEQ ID 625> is predicted to encode a protein having amino acid sequence <SEQ ID 626>:

1	MTFLQCMQGS	ADNKICAFAM	FVIRRFSEER	VQXAASMTF	TTLLALVPVL
51	TVMVAVASIF	PVFDRWSDSF	VSFNQITVP	QGADMVFDYI	DAFRDQANRL
101	TAIGSMVLVV	TSMLIRTID	NAFNRIWRVN	TQRFWMQFL	VYVWALLTFGP
151	LSLGVGISFM	VGSVQDSVL	SGAQWADAL	KTAARLAFT	LLWGLYRFF
201	PNRFVPAQQA	FVGLITAF	LETARFLTW	YMGNFYGRS	IYGAFAVFF
251	FLWLNLWLT	LVLGAVLTS	SLSYWQGEAF	RRGFDSRGR	DDVLKILL
301	DAAQKEGRTL	SVQEFRHHIN	MGYDELGELL	EKLARYGYI	SGRQGWVLT
351	GADSIELSEL	FKLFRVRLP	VERDHVNAV	DAVMTPCLQT	LNMTLAEFDA
401	QAKKQQS*				

Further work revealed the following gonococcal DNA sequence <SEQ ID 627>:

1	ATGACCTTTT	TACAACTGTG	GCAAGGTTTG	CGCGACAATA	AAATCTGTGC
51	ATTTCATGCG	TTCGTCATCC	GCGCTTTCAG	TGAAGACGCG	GTACCGCAGG
101	CAGCGGAGAG	CATGACGTTT	ACGACACTGC	TGGCACTCGT	CCCCGTACTG
151	ACCGTAATGG	TGCGCGGTGC	TTCGATTTC	CCGCTGTTCG	ACCGCTGGTC
201	GGATTGTTTC	GTCTCCTTGG	TCAACCAAA	CATTGTGCGC	CAGGCGCGCG
251	ATATGGTGTT	CGACTATATC	GACGCATTC	GGCATCAGCG	AAACCGGCTG
301	ACCGCATGCG	GCACGCTGAT	GCTGGCTGTA	ACCTGCTGTA	TGCTGATTCG
351	GACGATAGAC	AATGCTATCA	ACGCGACTGC	CGCGCTAAC	ACCGAACGCG
401	CTCGGATGAT	GCAGTTCCTC	GTTTATTGGG	CGTTGCTGAC	TTTCGGGCTC
451	TGTCTTTTGG	GTGTGGGCAT	TTCCTTTATG	GTCCGGTTCG	TTCACAGACT
501	CGTACTCTCC	TCCGGAGCGC	AACAATGGGC	GGACCGCTTG	AAGACGGCGG
551	CACGCGTGCG	TTTCATGACG	CTTTTGTCTG	GGGGCGTGTA	CCGCTCTCGT
601	CCCAACCGCT	TGCTGCCGCG	CCGCGAGGCG	TTTGTCCGAG	CTTTGATTAC
651	GGCATTTCTG	CTGAGAGAGC	CAGCTTTTCT	GTTTACCTGC	TATATGGGCA
701	ATTTCGACGG	CTACCGCTCG	ATTATCGGCG	CATTTCGCGC	CGTGGCGTTT
751	TTCTCTGCTG	GGTTAAACCT	CGTGTGAGAC	CTGGTCTTGG	GCGGGCGCGT
801	GCTGACTCTG	TGCTGCTCTT	ATTGGCAGGG	CGAGCGCTTC	CGCAGCGGAT
851	TGCACTCGCG	CGGACGGTTT	GACGACGTGT	TGAATACTCT	GCTGCTTCTG
901	GATGCGCGCG	AAAAAGAGGG	CCGAACCTTG	TCCGTTTCTG	AGTTTCAGACG
951	GCATATCAAT	ATGGGTTTAC	ATGAATTGGG	CGAGCTTTTG	GAAAAGCTGG
1001	CGCGGTACGG	CTATATCTAT	TCCGCGCAGC	AGGGCTGGGT	TTTGAARAAC
1051	GCGCGGGAAT	CGATTGAGTT	GAGCGAATCT	TTCAGACTCT	TGCTGTACCG
1101	CCCTGTGCTC	gtggaaAGGG	ATCATGTGAA	CCAAGCTGTC	gaTGGGTGTA
1151	TGAAGCGCTG	TTTGCAAGCT	TGACACATGA	CGCTGGCGGA	GTTTACGACT
1201	CAAGCGAATA	AACGACGACA	GTCTTGA		

This encodes a variant of ORF144ng, having the amino acid sequence <SEQ ID 628; ORF144ng-1>:

1	MTFLQRMQGL	ADNKICAFAM	FVIRRFSEER	VQXAASMTF	TTLLALVPVL
51	TVMVAVASIF	PVFDRWSDSF	VSFNQITVP	QGADMVFDYI	DAFRDQANRL
101	TAIGSMVLVV	TSMLIRTID	NAFNRIWRVN	TQRFWMQFL	VYVWALLTFGP
151	LSLGVGISFM	VGSVQDSVL	SGAQWADAL	KTAARLAFT	LLWGLYRFF
201	PNRFVPAQQA	FVGLITAF	LETARFLTW	YMGNFYGRS	IYGAFAVFF
251	FLWLNLWLT	LVLGAVLTS	SLSYWQGEAF	RRGFDSRGR	DDVLKILL
301	DAAQKEGRTL	SVQEFRHHIN	MGYDELGELL	EKLARYGYI	SGRQGWVLT

351 GADSIELSEL FKL FVYRPLP VERDHVNAV DAVMTPCLQT LNMTLAEFDA
401 QAKKQQQS*

ORF144ng-1 and ORF144-1 show 94.1% identity in 406 aa overlap:

```

5  orf144ng-1.pep  MTFLQRWQGLADNKAFAWVIRRFSEERVQAAASMTFTTLLALVPLVTVMVAVASIF
   orf144-1       MTFLQRWQGLADNKAFAWVIRRFSEERVQAAASMTFTTLLALVPLVTVMVAVASIF
10 orf144ng-1.pep  PVFDRWSDSFVSFVNQITIVPQADMVFDYIDAFRDQANRLTAIGSVMLVVTSLMLIRTID
   orf144-1       PVFDRWSDSFVSFVNQITIVPQADMVFDYIDAFRDQANRLTAIGSVMLVVTSLMLIRTID
15 orf144ng-1.pep  NAFNRIWRVNTQRFWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDSVLSGGAQWADAL
   orf144-1       NTFNRIWRVNSQRFWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDAALASGAPQWSGAL
20 orf144ng-1.pep  IYGAFAAVPFFLWMLNLWTLVLGGAULTSSLSYWGGEAFRRGFDSRGRFDDVLKILLLL
   orf144-1       IYGAFAAVPFFLWMLNLWTLVLGGAULTSSLSYWGGEAFRRGFDSRGRFDDVLKILLLL
25 orf144ng-1.pep  DAAQKEGRTLSVQEFRHHIMGYDELGELLEKLARYGIYSGRQGWLVKTGADSIELSEL
   orf144-1       DAAQKEGKALPVQEFRHHIMGYDELGELLEKLARHGIIYSGRQGWLVKTGADSIELSEL
   orf144ng-1.pep  FKL FVYRPLPVERDHVNAV DAVMTPCLQT LNMTLAEFDA QAKKQQQS
30 orf144-1       FKL FVYRPLPVERDHVNAV DAVMTPCLQT LNMTLAEFDA QAKKQQQS

```

On this basis of this analysis, including the identification of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

35 Example 75

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 629>:

```

1  ..AGACACGCC GCGCATCCG CATCGACACC GCCATCAACC CGGAACCTGGA
51 AGCCCTCGCC GAACACCTGC ACTACCAATG GCAGGGCTTC CTCGGCTCA
101 GCACCGCAT GCGTCAGGAA ATTTCGCCGC TCCTCATCGC GCTCGAACCG
40 151 ACCCGCCGCA ATGCGCTGGA TGCACAGAA CGCCACACCC TCGCCCAAG
201 CCGCTTGAA ACACGGGAAC ACGGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 630; ORF146>:

```

1  ..RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTDMRQE ISALVILLQR
51 TRRKWLDAHE RQHLRQSLLE TREHG*

```

45 Further work revealed the complete nucleotide sequence <SEQ ID 631>:

```

1  ATGAACACCT CGCAACGCCA CCGCTGTC AGCCGCTGCC TCAACTCTTA
51 CGAACGCTAC CGCTACCGCC GCTCATCCA CGCGCTCGG CTCGGCGGGG
101 CCGTCTCTGT CGCCACCGCC TCGCCCGGC TGCTCCACCT CCAACACGGC
50 151 GAGTGGATAG GGATGACCGT CTTCGTGTC CTCGGCATGC TCAGTTTCA
201 AGGGCGGATT TACTCCAAAG CGGTGGAACG TATGCTCGGC ACGGTCATCG
251 GGCTGGGGCG GGGTTTGGGC GTTTTATGCG TGAACACGA TATTATCCAC
301 GGCAACCTCC TCTTCTACCT CACCGTCGGC ACGGCAACGC CACTGGCCGG
351 CTGGGGCGGC GTGGGCAAAA ACGGCTACGT CCTATGCTG GCAGGGCTGA
401 CGATGTGTAT GCTCATCGGC GACAACGGCA GCGATGGCT CGACACGGGA
55 451 CTCATGCGCG CCATGAACGT CCTCATCGGC CGGGCCATCG CCATCGCCGC

```

501 CGCCAAACTG CTGCGCTGTA AATCCAACT GATGTGGCGT TTCACTGCTG
 551 CGCACACCT GCGCACTGC AGCAAAATGA TTGCCAAAT CACGCAACGC
 601 AGGCGCATGA CCGCGAAGC CCTCAGGAG AACATGGCGA AATGCGCCA
 651 AATCAACGCA CGCATGGTCA AAAGCCGCAG CCATCTCCGC GCCACATCGG
 701 GCGAAGCCGC CATCAGCCCC GCCATGATGG AAGCCATGCA GCACGCCAC
 751 CGTAAATTCG TCAACACCAC CGAGCTGCTC CTGACACGC CGCGCAAGCT
 801 GCAATCTCCC AAATCAACG CGAGCGAAAT CGGCTGCTT GACCGCACT
 851 TCACACTGCT CCAACCGAC CTGCAACAAA CGTGCCTCT TATCAACGGC
 901 AGACACGCCC GCGGCATCG CATCGACACC GCCATCAACC CGGAATGGA
 951 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
 1001 GCACCAATAT GCGTCAGGAA ATTTCCGCC TCGTCATCCT GCTGCAACGC
 1051 ACCGCGCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TCGCCAAAG
 1101 CTGCTTGAA ACACGGGAAC ACGGCTGA

This corresponds to the amino acid sequence <SEQ ID 632; ORF146-1>:

1 MNTSQNRNLV SRWLSYERY RYRRLIHAVR LGGVLFATA SARLLHLQHG
 51 EWIGMTVFVV LGMLQFOGAI YSKAVERMLG TVIGLGAGLG VLWLNQHYFH
 101 GNLLFYLTVG TASALAGWAA VGRNGIVFML AGLTMCMLIG DNGSEWLDG
 151 LMRAMNVLG AATAATAAKL LFLKSTLAWR EMLADNLAD SKMIEISNG
 201 RRMTRERLEE NRAMRQDNA RMVKSRLHA ATGSESRISF AMEAMQJH
 251 RRIVNTTELL LTAAKQLSP KLNGSEIRLL DRHFTLLQD LQQTVALING
 301 RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMREQ ISALVILLQR
 351 TRRWLDAHE RQLRQSLE TREHG*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF146 shows 98.6% identity over a 74aa overlap with an ORF (ORF146a) from strain A of *N. meningitidis*:

			10	20	30
orf146.pep			RHARRIRIDTAINPELEALAEHLHYQWQGF		
30 orf146a	KLNGSEIRLLDRHFTLLQDTDLQQTVALINGRHARRIRIDTAINPELEALAEHLHYQWQGF				
	280 290 300 310 320 330				
		40 50 60 70			
orf146.pep	LWLSTDMRQEISALVILLQRTTRRWLDAHERQHLRQSLETREHG				
35 orf146a	LWLSTNMREQEISALVILLQRTTRRWLDAHERQHLRQSLETREHSX				
	340 350 360 370				

The complete length ORF146a nucleotide sequence <SEQ ID 633> is:

40 1 ATGAACACCT CGCAACGCAA CCGCCTCGTC AGCGCTGGC TCAACTCCTA
 51 CGAACGCTAC CGCTACCGCC GCCTCATCCA CGCGCTCGG CTGCGGGGG
 101 CCGTCTCTGT CGCCACCGCC TCGCCCGGCG TGCTCCACT CCAACACGGC
 151 GAGTGGATGA GAGATGACGT CTTCGTGCTC CTGGCATGC TCCAGTTTCA
 201 AGGGGCGATT TACTCCAAGG CGGTGGAACG TATGCTCGG ACGGTATCG
 251 GCGTGGCGCG GGGTTTGGCG GTTTTATGCG TGAACAGCA TTATTTCCAC
 45 301 GGCAACCTCC TCTTCTACT CACGTCGCG ACGGCAAGCG CACTGCGCG
 351 CTGGGCGGCG GTCGGCAAAA ACGGCTACGT CCGTATGCT GCGGGCTCGA
 401 CGATGTGCAT GCTCATCGGC GACACGGCA GCGAATGGTT CGACAGGGCG
 451 CTGATCGCG CGATGACGT CTTATCGCG CGGGCATGC CACTGCGCG
 501 CGCCAACTG CTGCGCGTGA ATCCGCAACT GATGCGGT TTCAATCTTG
 55 551 CGCACACCT GACCGACTGC AGCAAAATGA TTGCCAAAT CACGCAACGC
 601 AGGCGCATGA CCGCGAAGC CCTCAGGAG AACATGGCGA AATGCGCCA
 651 AATCAACGCA CGCATGGTCA AAAGCCGCAG CCACCTCGCC GCCACATCGG
 701 GCGAAGCCGC CATCAGCCCC GCCATGATGG AAGCCATGCA GCACGCCAC
 751 CGTAAATTCG TCAACACCAC CGAGCTGCTC CTGACACGC CGCGCAAGCT
 55 801 GCAATCTCCC AAATCAACG CGAGCGAAAT CGGCTGCTT GACCGCACT
 851 TCACACTGCT CCAACCGAC CTGCAACAAA CGTGCCTCT TATCAACGGC
 901 AGACACGCCC GCGGCATCG CATCGACACC GCCATCAACC CGGAATGGA
 951 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
 1001 GCACCAATAT GCGTCAGGAA ATTTCCGCC TCGTCATCCT GCTGCAACGC
 60 1051 ACCGCGCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TCGCCAAAG
 1101 CTGCTTGAA ACACGGGAAC ACGTTCGA

This encodes a protein having amino acid sequence <SEQ ID 634>:

```

1  MNTSQNRNLV SRWLNSYERY RYRRLIHAVR LGGAVLFATA SARLLHLQHG
51 EWIGMTVFVV LGMLOFQGAI YSKAVERMLG TVIGLGAGLG VMLWLNQHYFH
101 GNLLFYLTVG TASALAGWAA VGKNGYVFM LAGLTMCLIG DNGSEWFDG
151 LMRAMNVLIG AATAIAAAKL LPLKSTLMWR FMLADNLTDC SKMIAEISNG
201 RRMTRERLEE NMAKMRQINA RMVKSRSHLA ATSGESRISP AMMEAMQHAH
251 RKIVNTTELL LTAAKLQSP KINGSEIRLL DRHFTLLQTD LQQTVALING
301 RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMROE ISALVILLOR
351 TRRWLDAHE RQHLRQSLE TREHS*

10  ORF146a and ORF146-1 show 99.5% identity in 374 aa overlap:

    orf146a.pep  MNTSQNRNLVSRWLNSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFVV
    orf146-1      MNTSQNRNLVSRWLNSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFVV
15  orf146a.pep  LGMLOFQGAIYSKAVERMLGTVIGLGAGLGVLWLNQHYFHNLLFYLTVGTSALAGWAA
    orf146-1      LGMLOFQGAIYSKAVERMLGTVIGLGAGLGVLWLNQHYFHNLLFYLTVGTSALAGWAA
20  orf146a.pep  VGKNGYVFM LAGLTMCLIGDNGSEWFDGSLMRAMNVLIGAAIAAAAKLLPLKSTLMWR
    orf146-1      VGKNGYVFM LAGLTMCLIGDNGSEWFDGSLMRAMNVLIGAAIAAAAKLLPLKSTLMWR
25  orf146a.pep  FMLADNLTDCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP
    orf146-1      FMLADNLTDCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP
30  orf146a.pep  AMMEAMQHAHRKIVNTTELLLTAAKLQSPKINGSEIRLLDRHFTLLQTDLQQTVALING
    orf146-1      AMMEAMQHAHRKIVNTTELLLTAAKLQSPKINGSEIRLLDRHFTLLQTDLQQTVALING
    orf146a.pep  RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMROEISALVILLQRTTRRWLDAHE
    orf146-1      RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMROEISALVILLQRTTRRWLDAHE
35  orf146a.pep  RQHLRQSLE TREHSX
    orf146-1      RQHLRQSLE TREHG

```

Homology with a predicted ORF from *N.gonorrhoeae*

40 ORF146 shows 97.3% identity over a 75aa overlap with a predicted ORF (ORF146ng) from *N.gonorrhoeae*:

```

    orf146.pep  RHHARRIRIDTAINPELEALAEHLHYQWQGF  30
    orf146ng    KINGSEIRLLDRHFTLLQTDLQQTAAALNGRHARRIRIDTAINPELEALAEHLHYQWQGF  364
    orf146.pep  LWLSTNMROEISALVILLQRTTRRWLDAHERQHLRQSLE TREHG  75
    orf146ng    LWLSTNMROEISALVILLQRTTRRWLDAHERQHLRQSLE TREHG  409

```

An ORF146ng nucleotide sequence <SEQ ID 635> was predicted to encode a protein having amino acid sequence <SEQ ID 636>:

```

1  MSGVRFPSPA PIPSTDPPSG SLCTFTFPLQ TASDMNSQR KRLSGRWLNS
51 YERYRHRRLI HAVRLGGTVL FATALARLLH LQHGGEWIGMT VFVVLGMLQF
101 QGAIYSNAVE RMLGTVIGLG AGLGVLWLNQ HYFHGNLLFY LTIGTASALA
151 GWAAVGKNGY VFMLAGLTM LIGDNGSEW LDSGLMRAMN VLIGAATAIA
201 AAKLLPLKST LMRFMLADN LADCSKMIAE ISNGRRMTR ELEGNMVKMR
251 QNARMVKSR SHLAATSGE RISPMMMEAM QHAHRKIVNT TELLITTAAK
301 LQSPKLGSE IRLDRHFTL LQTDLQQTAA LINGRHARRI RIDTAINPEL
351 EALAEHLEYQ WQGF LWLSTN MROEISALVI PLQRTTRRWL DAHERQHLRQ
401 SLE TREHG*

```

60 Further work revealed the following gonococcal DNA sequence <SEQ ID 637>:

```

      1 ATGAACCTCT CGCAACGCAA ACGCCTTTCC GgcCGTGGC TCAACTCCTA
      51 CGCAACCTAG CGCAACGCGC GCTCATACG TCGCCGCGGG CTGGCGCGAA
     101 cegTCCTGTT CGCCACCGCA CTCGCCCGgc LACTCCACCT CCAacacggc
     151 qAATGGATAG GGAAtgACGT CTTCTGTGCT CTCGGATGCG TCCAGTTCCA
     201 AGGCGcgatT tActccaacg cggTgGAacg tatGTCcggt acggtcatcg
     251 ggcTgGGCGC GGGTtTGGgc gTTTtATGgc TGAACCGAGA TtAttccacg
     301 gqcaacCTcc tcttctacct gaccatcggc acgGcaagcg cactggccgg
     351 ctGGCGCGCG GTCGGCAAAA acggctacgt cccatgtcg GCGGGGctgA
     401 CGATGTGCAT gctcatcggc gACAACGGCA GCGAATGGCT GCACAGCGCG
     451 CTGATGGCGG CGATGAACGT CCTCATCGCG GCGCGCATCG CCATTGCCCG
     501 CGCCAAACtG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTtG
     551 CGCAACAACCT GGCCGACTCG AGCAAAATGA TTGCGGAAT CAGCAACGCG
     601 AGGCGTATGA CGCGCGAAGC TTTGGAGCAG AATATGGTCA AAATGCGCCA
     651 AATCAACGCA CGCATGGTCA AAGCCGCGAG CCACCTGCCG GCCACATCGG
     701 GCGAAAGCCG CATCAGCCCC TCCATGATGG AAGCCTATGCA GCACGCCCACT
     751 CGCAAAATCG TCAACACCAc CGAGCTGCTC CTGACACCGC CGCCAAAGCT
     801 GCAATCTGCC AACTCAACG CGAGCGAATC CGCGCTGCTC GACCGCCACT
     851 TCACACTGCT CCAACCGCAc CTGCAACAA CGCCGCCCT CATCAAGCGG
     901 AGCACGCGCC GCCCATCCCG CATTGACACG GCATCAACc CGAACTGGA
     951 AGCCCTGCCG GAACACCTCC ACTACAATG GCAGGGTTC CTCTGGCTCA
    1001 CGACCAATAT GCGTCAGGAA ATTTCCGCC TCCTCATCCT GTCGCAACCG
    1051 ACCCGCGCCA AATGGCTGGA TGCCACGAA CGCCAACACC TGCCGCAAAG
    1101 CTGCTTGAA ACACGGGAAC ACGGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 638; ORF146ng-1>:

```

     25      1 MNSSQRKRLS GRWLSNYERY RHRRLIHAVR LGGTVLFATA LARLLHLQHG
     51 EWIGMTVFVV LGMLQFGAI YSNAVERMLG TVIGLGAGLG VLWLNQHYFH
    101 GNLLFYLTIG TASALAGWAA VGKNGYVFM LAGLTMCLIG DNGSEWLDG
    151 LMRAMNVLIG AATAIAAAKL LPLKSTLMWR FMLADNLDC SKMIAEISNG
    201 RMRTRERLEQ NMVVKRQINA RMVKSRSHLA ATSGESRISP SMMEAMQHAH
    251 RKIVNTTELL LTTAAKLQSP KLNGSEIRLL DRHFTLLQTD LQQTAAALING
    301 RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMROQE ISALVILLOR
    351 TRRWLDAAHE RQHLRQSLLE TREHG*

```

ORF146ng-1 and ORF146-1 show 96.5% identity in 375 aa overlap

```

    35 orf146-1.pep MNTSQNRNLSRWLSNYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGIEWIGMTVFVV
      orf146ng-1 MNSSQRKRLSGRWLSNYERYRHRRLIHAVRLGGTVLFATALARLLHLQHGIEWIGMTVFVV
      orf146-1.pep LGMLQFGAIYSKAVERMLGTVIGLGAGLGVLWLNQHYFHNLLFYLTIVGTASALAGWAA
      orf146ng-1 LGMLQFGAIYSNAVERMLGTVIGLGAGLGVLWLNQHYFHNLLFYLTIGTASALAGWAA
    40 orf146-1.pep VGKNGYVFM LAGLTMCLIGDNGSEWLDGSLMRAMNVLIGAAIAIAAAKLPLKSTLMWR
      orf146ng-1 VGKNGYVFM LAGLTMCLIGDNGSEWLDGSLMRAMNVLIGAAIAIAAAKLPLKSTLMWR
    45 orf146-1.pep FMLADNLDCSKMIAEISNGRMRTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP
      orf146ng-1 FMLADNLDCSKMIAEISNGRMRTRERLEQNMVVKRQINARMVKSRSHLAATSGESRISP
    50 orf146-1.pep AMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNGSEIRLLDRHFTLLQTDLQQTVALING
      orf146ng-1 SMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNGSEIRLLDRHFTLLQTDLQQTAAALING
    55 orf146-1.pep RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMROQEISALVILLQTRRWKLDAAHE
      orf146ng-1 RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMROQEISALVILLQTRRWKLDAAHE
      orf146-1.pep RQHLRQSLLETREHGX
    60 orf146ng-1 RQHLRQSLLETREHGX

```

Furthermore, ORF146ng-1 shows homology with a hypothetical *E.coli* protein:

```

    65 sp|P33011|YEEA ECOLI HYPOTHETICAL 40.0 KD PROTEIN IN COBU-SBMC INTERGENIC REGION
      >gi|1736674|gnl|PIDD|d1016553 (D90838) ORF_ID:0348#20; similar to [SwissProt
      Accession Number: P33011] [Escherichia coli] >gi|1736682|gnl|PID|d1016560 (D90839)
      ORF_ID:0348#20; similar to [SwissProt Accession Number: P33011] [Escherichia coli]

```

-355-

>gill1788318 (AE000292) f352; 100% identical to fragment YEEA ECOLI SW: P33011 but has 203 additional C-terminal residues [Escherichia coli] Length = 352
Score = 109 bits (271), Expect = 2e-23
Identities = 89/347 (25%), Positives = 150/347 (42%), Gaps = 21/347 (6%)

5 Query: 20 YRHRRLIHAVRLGSTVLFATALARLLHLQHGEWIGMTVFVVLGMLQFQGAISYNAVERML 79
YRH R++H R+ L + RL + W +T+ V++G + F G + A ER+
Sbjct: 15 YRHYRIVHGTRVALAFLTLFLIIRLFTYPESTWPLVTMVVIMGPISFWGVXVVPRAFERIG 74

10 Query: 80 GTVIGLAGLGVWLNLQHYFHGNLLFYLTGTASALAGWAAVKGNGVVPMLAGLTMCLM 139
GT+G GL L L L + A L GW A+GK Y +L G+T+ ++
Sbjct: 75 GTVLGSLGLIALQLE---LISLPLMLVWCNAAMFLOGMLALGKYPYGLLLIGVTLAIVV 131

15 Query: 140 GDNGSEWLDGLMRAMNVLIGXXXXXXXXKLLPLKSTIMWRFLMADNLADCSKMAIEISN 199
G E +D+ L R+ +V++G + P ++ + WR LA +L + ++ +
Sbjct: 132 GSPTGE-IDTALWRSQDVILGSLAMLETGIWQRAFIHRIQLAKSLTEYNRVYQSAFS 190

Query: 200 GRMTRERLEQNMVMMKQINARMMVKSRSHLAATSGESRISPSMMEAMQHAHRKIVNXXXX 259
+ R RLE ++ K+ VK R+ A S ERI S+ E +Q +R +V
Sbjct: 191 PNLLERPRLESHLQKLL---TDAVKMRGLIAPASKEPTRIPKSYEGIGTINRNLVCMLEL 247

20 Query: 260 XXXXXXXXQSPK---LNGSEIRLLDRHFXXXXXXXXXAALINGRHARRIRIDPAINPEL 316
+ LN ++R D AL G +N +
Sbjct: 248 QINAWYATRPSHFVLLNAQKLR--DTQHMQQILLSLVHALYEGNPQVPEANTEKLNDAV 305

25 Query: 317 EALAHL--HYQWQ-----GFLWLSNMNRQEISALVILLQRTRRK 354
E L + L H+ + G++WL+ ++ L L+ R RK
Sbjct: 306 EELRQLLNHHHDLKVVETPIYGYVWLNMETAQLELLSNLICRALRK 352

On the basis of this analysis, including the identification of several transmembrane domains in the
gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and
their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 76

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 639>

35 1 ..GCCGAAGACA CGCGCGTTAC CGCAGAGCTT TTGAGCGCGT ACGSCATTCA
51 GCGCAAACTC GTACAGTGTG CGGACACAAA CGAACGGCAG ATGCGCGACA
101 AGATTGTGCG CTATCTTTCA GAGCGCATGG TTGTGCGACA GGTTCGGCAT
151 GCGGATAGCG CGCGCGGTGG CGACCCGCGCG CGGAAACTCG CCGCGCGCAT
201 CGGTGAGGCC GGGTTTAAAG TCGTTCGCCGT CGTGGGCGCA AC.GCGGTGA
40 251 TGCGCGGCTTT GAGCGTGGCC GGTGTGGAAG GATCCGATT TTTATTTCAAC
301 GGTTTGTGAC CGCGGAATC GSGAGAAGCG AGGAACTGT TTGCCAAATG
351 GGTGCGGGCG GCGTTTCTCA TCGTCTGTT TGAAGACCG CACGCGATCG
401 GTGCAAGCGT TGCGGATATG CGGGAAGTGT TCCCGAAGC CGGATTAATG
451 CTGGGCGGCG AATATTACGAA AACGTTTGAA ACGTCTTAA CGGCGACGGT
501 TGGGGAATTT CAGACGGCAT TGTCTGCGGA CGGCGACCAA TCGCGCGGGG
45 551 AGATGTTGTT GGTGCTTTAT CCGGCGCAG ATGAAACAA CAAAGGCTTG
601 TCCGAGTCCG CGCAAAACAT CATGAAATC CTCACAGCG AGCTCGCGAG
651 CAACACGGCG CGGACGCTTG CTGCCAAAT CACGGGCGAG GGAAGAAAG
701 CTTGTGACGA T..

This corresponds to the amino acid sequence <SEQ ID 640; ORF147>:

50 1 ..AEDTRVTAQL LSAYGIQGL VSVREHNERQ MADKIVGYLS DGMVVAQVSD
51 AGTPAVCDPG AKLAREVREA GFKVFPVPGA XAVMAALSVA GVEGSDFYFN
101 GFVPKPSGER RKLFAKWRVA AFPVIMFETP HRIGAAALADM AELFFERRIM
151 LAIREITKFE TFLSGTVGEI QTALSADGD SRGEMVLVLY PAQDEKIEGL
201 SESAQINIKI LTAELPTQA AELAAKITGE GKALYD..

55 Further work revealed the complete nucleotide sequence <SEQ ID 641>:

1 ATGTTTCAGA AACATTTCGA GAAAGCTCC CACAGCGTCG TCGGAGGGAC
51 ATTATAOCTG GTTGCACGCG CCATCGGCAA TTGTCGGGAC ATTACCTCGT
101 GCGCTTTGGC GGTATTGCAA AAGCGGACA TCATCTGTGC CGAAGACAG
151 CGCGTTACCG CACAGCTTTT GAGCGCGTAC GGCATTACAG GCAAACTCGT

```

201 CAGTGTGCGC GAACACAACG AACGCGAGT GCGGCAAG ATTGTGCGCT
251 ATCTTTTCAG CCGATGGT GTGACACAG TTTCGATCG GGTACGCGC
301 CGCGTTCGCG ACCCGGCGCG GAAACTCGCC CGCGCGCTGC GTGAGGCGCG
351 GTTTAAAGTC GTTCCCGTCG TGGGCGCAAG CGCGCTGATG GCGGCTTTGA
401 GCGTGGCCGG TGTGAAGGA TCCGATTTTT ATTTCAACGG TTTTGTACCG
451 CCGAAATCGG GAGAACGCGA GAAACTGTTT GCCAAATGCG TCGCGGCGCG
501 GTTTCCTATC GTCATGTTTG AAACCGCCGA CCGCATCGGT GCGACGCTTG
551 CCGATATGGC GGAACGTGTC CCGGAACGCC GATTAATGCT GCGCGCGCAA
601 ATTACGAAAA CGTTTGAACG GTTCTTAAGC GGCACGGTTG GGGAAATCCA
651 GACGCGCATTG TCTGCGCGAC GCAACCAATC GCGCGCGCAG ATGTGTTGG
701 TGCTTTATCC GCGCGCAGGT GAAAAACACG AAGGCTTGTG CGAGTCCGCG
751 CAAAAACATCA TGAATACTCT CACAGCGCAG CTGCCGACCA AACAGCGCGC
801 GGAGCTTGCT GCCAAATCA CCGGCGAGGG AAAGAAAGCT TTGTACGATC
851 TGGCTCTGTC TTGAAAAAAC AATAG

```

15 This corresponds to the amino acid sequence <SEQ ID 642; ORF147-1>:

```

1 MFQKHIQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT
51 RVTAQLLSAY GIQGLVSVR EHNERQMAK IVGYLSDGMV VAQVSDAGTP
101 AVCDGAKLA RRVREAGFKV VPVVGASAVM AALSVAQVEG SDFYFNFGVP
151 PKSGERRKLF AKWVRAAFP VMFETPHRIG ATLADMAELF PERRLMARE
201 ITKTFETFLS GTVGEIQTAL SADGQSBGE MVLVLYPAQD EKHEGLSESA
251 QNIMKILTAE LPTKQAAELA AKITGEGKKA LYDLALSWGN K*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical protein ORF286 of *E. coli* (accession number U18997)

ORF147 and *E. coli* ORF286 protein show 36% aa identity in 237aa overlap:

```

25 Orf147: 1 AEDTRVTAQLLSAYGIQGLVSVREHNERQMAKIVGYLSDGMVVAQVSDAGTPAVCDPG 60
AEDTR T LL +GI +L ++ +HNE+Q A++ ++ L +G +A VSDAGTP + DEG
Orf286: 43 AEDTRHTGLLLQHFGINARLFAIHDHNEQQAETLLAKLQEQGNITLVSDAGTPLINDPG 102

30 Orf147: 61 AKLARVRREXXXXXXXXXXXXXXXXXXGSDFYFNFGVPPKSGERRKLFKAVWRA 120
L R RE F + GF+P KS RR
Orf286: 103 YHLVRTCREAGIRVVPLPGCAAITALSAGLPSDRFCYEGFLPAKSGKRDRALKAIKAE 162

Orf147: 121 APFIVMFETPHRIGAAALADMAELFPERR-LMLAREITKTFETFLSGTVEIQTALSADGD 179
++ +E+ HR+ +L D+ + E R ++LARE+TKT+ET VGE+ + D +
Orf286: 163 PRTLIIYESTHRLDLSLEDIVVLGSESRVYLARELTKTWETIHGAPVGLLAWKEDEN 222

Orf147: 180 QSRGEMVLVLYPAQDEKHEGLSESAQNIMKILTAE LPTKQAAELA AKITGEGKKALY 236
+ +GEMVL++ + E L A + +L AELP K+AA LAA+I G K ALY
Orf286: 223 RRGEMVLIV-EGHKAQEDLPADALRTLALLOAELPLKKAALAAEIHGVKKNALY 278

```

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF147 shows 96.6% identity over a 237aa overlap with ORF75a from strain A of *N. meningitidis*:

```

45 orf147.pep AEDTRVTAQLLSAYGIQGLVSVREHNERQ 10 20 30
|||
orf75a TLYVATPIGNLADITLRALAVLQKADII CAEDTRVTAQLLSAYGIQGLVSVREHNERQ
20 30 40 50 60 70
50 orf147.pep MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKIARRRVREAGFKVVPVVGASAVMAALSV
|||
orf75a MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKIARRRVREAGFKVVPVVGASAVMAALSV
80 90 100 110 120 130
55 orf147.pep 100 110 120 130 140 150
GVGSDFYFNFGVPPKSGERRKLFKAWVRAAFPVIMFETPHRIGAAALADMAELFPERRLM
|||
orf75a GVGSDFYFNFGVPPKSGERRKLFKAWVRAFPVIMFETPHRIGATLADMAELFPERRLM
140 150 160 170 180 190
60 orf147.pep 160 170 180 190 200 210
LAREITKTFETFLSGTVEIQTALSADGQSRGEMVLVLYPAQDEKHEGLSESAQNIMKI

```


	orf75a	 LAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEKHEGLSESQAQNMKI 200 210 220 230 240 250
5	orf147.pep	220 230 LTAEFLPTKQAAELAAKITGEGKKALYD LTAEFLPTKQAAELAAKITGEGKKALYDLALSWKNKX
	orf75a	260 270 280 290
10	ORF147a is identical to ORF75a, which includes aa 56-292 of ORF75.	
	<u>Homology with a predicted ORF from <i>N.gonorrhoeae</i></u>	
	ORF147 shows 94.1% identity over a 237aa overlap with a predicted ORF (ORF147ng) from <i>N. gonorrhoeae</i> :	
15	orf147.pep	AEDTRVTAQLLSAYGIQGLVSVREHNERQ 30
	orf147ng	TLVVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQGLVSVREHNERQ 85
	orf147.pep	MADKIVGYLSDGMVVAQVSDAGTFPAVCDPGAKLARRVREAGFKVVPVVGASAVMAALSVA 90
20	orf147ng	MADKIVGYLSDGLVVAQVSDAGTFPAVCDPGAKLARRVREAGFKVVPVVGASAVMAALSVA 145
	orf147.pep	GVEGSDFYFNGFVPPKSGERRKLFAKWVRAAFPVVMFETPHRIGATLADMAELFFERRIM 150
25	orf147ng	GVAESDFYFNGFVPPKSGERRKLFAKWVRAAFPVVMFETPHRIGATLADMAELFFERRIM 205
	orf147.pep	LAREITKTFETFLSGTVGEIQTALADGNQSRGEMVLVLYPAQDEKHEGLSESQAQNMKI 210
	orf147ng	LAREITKTFETFLSGTVGEIQTALADGNQSRGEMVLVLYPAQDEKHEGLSESQAQNMKI 265
30	orf147.pep	LTAEFLPTKQAAELAAKITGEGKKALYD 237
	orf147ng	LAELPTKQAAELAAKITGEGKKALYDLALSWKNK 300

An ORF147ng nucleotide sequence <SEQ ID 643> was predicted to encode a protein having amino acid sequence <SEQ ID 644>:

35	1 MSVFQTAFFM FQKHLQKASD SVVGGTLYVV ATPIGNLADI TLRALAVLQK 51 ADIICAEDTR VTAQLLSAYG IQGRLVSVRE HNERQMAQDV IGFLSDGLV 101 AQVSDAGTFA VCDPGAKLAR VRREAGFKVV PVVGASAVMA ALSVAGVAES 151 DFYFNGFVFP KSGERRKLFA KWVRAAFPVV MFETPHRIGA TLADMAELFF 201 ERRMLAREI TKTFTFLLSG TVGEIQTALA ADGNQSRGEM VLVLYPAQDE 40 251 KHEGLSESAQ NAMKILAAEL PTKQAAELAA KITGEGKKAL YDLALSWKNK 301 *
----	---

Further work revealed the following gonococcal DNA sequence <SEQ ID 645>:

	1 ATGTTTCAGA AACACTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC 51 ATTATACGTG GTTGCCACGC CCATCGGGCAA TTTGGCAGAC ATTACCCCTGC 45 101 GCGCTTTGGC GGTATTGCAA AAGCGCGACA TCATTTTGTC CGAAGACACG 151 CGGCTTACTG CGCAGCTTT GAGCGCGTAC GGCATTGAGC GCAGGTTGTT 201 CAGTGTGCGC GAACACAACG AGCGGCGAGT GCGGACAGAG GTAATCGGTT 251 TCCTTTTCAGA CGGCTGTGTT GTGGCGCAGG TTTCGATGTC GGGTACGCGC 301 GCGGTGTGCG ACCCGGGGCG GAAACTCGCC CGCCGCGTGC GCGAAGCAGG 50 351 GTTCAAGTCT GTTCCGCTCG TGGCGCCAGC CGCGGTAAATG CGCGGTTTGA 401 GTGTCGCGCG TGTGGCGGAA TCCGATTGTT ATTTCAACGG TTTTGTACCG 451 CGGAATTCGG CGCAACGTAG GAAATGTGTT GCCAATGSG TCGCGGCGGC 501 ATTTCTCTGTG GTCATGTGTT AAAOCCGCCA CGCAATCGGG GCAACGCTGT 55 551 CGGATATGGC GGAATGTGTT CCGCAACGCC GTCTGATGCT GCGCGCGGAA 601 ATCAGCAAAA CGTTTGAACG GTTCTTAAGC GGCACGGTGT GGGGAATTTCA 651 GACGCGCAATG GCGGCGGACG GCAACCAATC GCGCGCGGAG ATGCTGTTGG 701 TGCTTTATCC GGCAGGAGAT GAAACACACG AAGCGTTGTC CGAGTCTGCG 751 CAAATGCGA TGAAATCCT TCGCGCGGAG CTGCGGACCA AGCAGGCGGC 801 GGAGCTTGCC GCAAGATTA CAGTGAGGG CAAAGAGCT TGTGACGATT 60 851 TGGCACTGTC GTGGAAACAA AATGA
--	--

This corresponds to the amino acid sequence <SEQ ID 646; ORF147ng-1>:

```

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEET
51 RVTAQLLSAY GIQGRIVSVR EHNERQMAK VIGFLSDGLV VAQVSDAGTP
101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAAGVAE SDFYFNGFVP
151 PKSGERRKLF AKWVRAAPFV VMEETPHRIG ATLDMAELEF PERRIMLAER
201 ITKTFETFLS GTVGEIQFAL AADGNQSRGE MVLVLYPAQD EKHEGLSESA
251 QNAMKILAAE LPTQAAELA AKITGEGKKA LYDLALSWKN K*
```

ORF147ng shows homology to a hypothetical *E. coli* protein:

```

10 sp|P45528|YRAL_ECOLI HYPOTHETICAL 31.3 KD PROTEIN IN AGAI-MTR INTERGENIC REGION
   (F286)
   >gi|606086 (U18997) ORF_f286 [Escherichia coli]
   >gi|1789535 (AE000395) hypothetical 31.3 kd protein in agai-mtr intergenic region
   [Escherichia coli] Length = 286
   Score = 218 bits (550), Expect = 3e-56
15 Identities = 128/284 (45%), Positives = 171/284 (60%), Gaps = 4/284 (1%)

Query: 4 KHLQKASDSVVGGTLYVATPIGNLADITLRALAVLQKADIICAEETRVTAQLLSAYGIQ 63
      K Q A +S G LY+V TPIGNLADIT RAL VLQ D+I AEDTR T LL +GI
Sbjct: 2 KQKQASDANSQ--GQLXIVPTPIGNLADITQRALEVLQAVDLIAEDTRHTGLLLQHFNG 59

20 Query: 64 GRIVSVREHNERQMAKDVIGFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPV 123
      RL ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG L R REAG +VVP+
Sbjct: 60 ARLFALHDHNEQQAETLLAKLQEQGNIALVSDAGTPLINDPGYHLVRTCREAGIRVPL 119

25 Query: 124 VGASAVMAALSAGVVAESDFYFNGFVPPKSGERRKLFKAKWVRAAPFVVMETPHRIGATL 183
      G A + ALS AG+ F +GF+P KS RR ++ +E+ HR+ +L
Sbjct: 120 PGCAAITALSAGLPSDFCYEGFLPAKSKGRDALKAEAEPTLFIYESTHRLDLSL 179

30 Query: 184 ADMAELFPERR-IMLAEREITKTFETFLSGTVEIQTALAADGNQSRGENVLVLYPAQDEK 242
      D+ + E R ++LARE+TKT+ET VGE+ + D N+ +GENLV++ +
Sbjct: 180 EDLVAVLGESRVVVLARELTKTWETIHGAPVGEALLAWVKEDENRRKGMVLV-EGHKAQ 238

Query: 243 HEGLSESAQNAMKILAAELPTQAAELAAKITGEGKALYDLAL 286
      E L A + +L AELP K+AA LAA+I G K ALY AL
35 Sbjct: 239 EEDLPADALRTLALLQAEPLPKKAALAAEIHGVKNALYKYAL 282
```

Based on the computer analysis and the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that these proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 77

40 The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 647>

```

1 ATGAAACCAA CGACAAACG GACAAACGAA ACACACCCGA AAGCCCGCAA
51 AACCGGTGCG ATCCGCTTCT C.GCTGCTTA CTTAGCCATA TGCCTGTGCT
101 TCGGCATCTCT TCCCGAAGCC TGGGCGGGAC ACACATATTAT CGGCATCAAC
151 TACCACTACT ATCCGCACTT TCCCGAAGAT AAGGCAAGGT TTGCGATGCG
45 201 GGCCAAGAT ATTCAGGTTT ACACAAATAA AGGGGATGTG GTGCGCAART
251 CAATGACAAA AGCCCCGATG ATTGATTTTT CTGTGGTGTC GCGTAAACGC
301 GTGGGGCGAT TGGTGGGCGT ATCAATATAT TGTGAGCGTG GCACATAACG
351 GCGGCTATAA CAACGTTGAT TTTGTGCGG AAGGAAK. AA TATCCG. GAT
401 CAACACGw TACTACTATAA AATTGTGAAA CGGAATATATT ATAAGCAGG
50 451 GACTAAAGGC CATCCTTATG GCGGCGGATTA TCATATGCGC CGTTTGCATA
501 AATwTGTAC AGATGCAGAA CCTGTGAAA TGACCATGTA TATGGATGGG
551 CGGAATATA TCGATCAAAA TAATTACCCT GACCGTGTC GTATTGGGGC
601 AGGCAGGCAA TATTGGCAGT CTGATGAAGA TGAGCCAAT AACCCGAA
651 GTTCATATCA TATTGCAAGT .....
701 ..... GGCTC ACCAATGTT ATCTATGATG CCGAAAAGCA
751 AAGTGGTTA ATTAATGGGG TATTGCAAA CCGGCAACCC TATATAGGAA
801 AAGACGATGG CTTCAGCTG GTTCGTAAG ATTGGTTCTA TGATGAATC
851 TTTGCTGGAG ATACCATTC AGTATTCTAC GAACCAAGT AAAATGGGAA
901 ATACTCTTTT AACGACGATA ATAATGGCAC AGGAAAAGAA ATATGCAAA
```

951 ATGAACACAA TTCTCTGCTT AATGATTAA AAACAGGAAC CGTCAATTG
 1001 TTTAAGCTTT GTTTATCCGA GACAGCARGA GAACCTGTTT ATCATGCTGT
 1051 AGGTGGTGCTC AACAGTTATC GACCAGACTC GATAATGGGA GAAATATATT
 1101 CCTTTATTGA CGAAGGAAAA GGGGAATTGA TACTTACCAG CAACATCAAT
 1151 CAAGGTGCTG GAGGATTATA TTTCCAAGGA GATTTTACGG TCTCGCCTGA
 1201 AAATAACGAA ACTTGGCAAG GCGCGGGCGT TCATATCAGT GAAGACAGTA
 1251 CGGTACTTGT GAAAGTAAAC GGGCTGGCAA ACGACGCGCT GTCCAAAATC
 1301 GGCAAAGGCA CGCTG.....
 //
 10 2101GATAAAG
 2151 TGACTGCTTC ATTGACTAAG ACCGACATCA GCGGCAATGT CGATCTTGCC
 2201 GATCAGCCTC ATTTAAATCT CACAGGGCTT GCCACACTCA ACGGCAATCT
 2251 TAGTGCAATT GGGGATACAC GTTATACAGT CAGCCACAAC GCCACCCAAA
 15 2301 ACGGCAACCK TAGCCTCGTG G. sAATGCC AAGCAACATT TAAATCAAGC
 2351 ACATTAAGCG GCAACACATC GGCTTCGGGC AATGCTTCAT TTAATCTAAG
 2401 CGACCAAGCG GTACAAAACG CAGTCTGAC GCTTTCGGCG AAGCGTAAGG
 2451 CAAGCTTAAG CCATTCCGCA CTCACAGGTA ATGTCTCCCT ACGCGATAAG
 2501 CGAGTATTC ATTTTGAAGC GACGCGCTT ACGCGACAAA TCAGCGCGCG
 2551 CAAGGATACG GCATTACACT TAAAGACAG CSAATGGAGC CTGCGCTCAG
 20 2601 Ga rCGGAATT AGGCAATTTA AACCTTGACA ACGCCACCAT TACACTCAAT
 2651 TCGCGCTTAT GCCACGATCG GGCAGGGGCG CAACCGCGCA GTGCGACAGA
 2701 TGGCGCGCGC CGCGCTTCGC GCGCTTCGCG CGGTTCCTCA TTAThCGTTA
 2751 CACGCGCAAC TTCGGTAGAA TCCCGTTTCA ACACGCTGAC GGTAAACGGC
 2801 AAATTGAACG GTCAGGGAAC ATTCCGCTTT ATGTGCGAAG TCTTCGGCTGA
 25 2851 CCGCAGCGAC AAATTGAAGC TGGCGGAAGG TTCGGAAGGC ACTTACACTC
 2901 TGGCGGTCAA CAATACGCGC AAGCAACCTG CAAGCTCGGA ACAATTGACG
 2951 GTAGTGGRAG GAAAGACAAA CAACCGCTG TCGAAAACCC TTAATTTTAC
 3001 CTGCAAAACG GAACACGTG ATGACGGGCG GTGG.....
 //
 30 3551TTAGAC CGOGTATTGG CCGAAGACCG
 3601 CGGCAACGCC GTTTGGACAA GGGGCATCCG GGACACCAAA CACTACCGTT
 3651 CGCAAGATTT CGCGCGCTAC CGCCAAACAA CGGACTCTGG CCAATACGGT
 3701 ATGCAAGAAA ACTCTGGCAG CGGGCGGCTC GGCATCTGTT TTTCGACAAA
 3751 CGGCAAGAAA AACACTTCG ACAGCGCAT CGGCAACTCG CACGCGCTTG
 35 3801 CCCACGCGCG CGTTTTCGGG CAATACGGCA TCGCACGGTT CTACATCGCG
 3851 ATCAGGCGCG GCGCGGGTT TTAGCAGCGG CAGCCTTTCA GACGCGATCG
 3901 GAGmAAAwT CGCGCGCGC GTGCTGCATT ACGCATTCGA GCGCAGAtAC
 3951 CGCGCGGgtt tCggCGgAtt GGCATCGAA CCGCACATCG GCGCAACGcg
 40 4001 ctATTTCGTC CAAAAGCGG ATTACCGCTA CGAAAACGTC AATATCGCCA
 4051 CCCCOCGGCT TGCATTCAAC GCGTACCGCG CGGGCATTAA GGCAGATTAT
 4101 TCATTCAAAC CGGCGCAACA CATTTCACAT ACGCCTTATT TGAGCCTGTC
 4151 CTATACCGAT CGCGCTTCGG GCAAAGTCG AACACGCGTC AATACGCGCG
 4201 TATTGGCTCA GGATTTCGGC AARACCCGCA GTGCGGAATG GggCGTAAAC
 4251 GCGGAAATCA AAGGTTTCAAC GCTGTCCCTC CACGCTGCGC CGGCCAAAGG
 45 4301 CCGCAACTCG GAAGCGCAAC ACAGCGCGGG CATCAAAATTA GGCTACCGCT
 4351 GGTAA...
 //
 701DKVTAS LTKTDISGNV DLADHAHLNL TGLATLNGNL
 751 SANGDTRYTV SHNATQNGNX SLVXNAQATF NQATLNGNTS ASGNASFNLS
 801 DHAVQNGSLT LSGNAKANVS HSLANGNVSL ADKAVFFIFES SRTFGQ136G
 851 KDTALHLKDS EWTLP5GKEL GNLNLDNATI TLNSAYRHDH AGAQTGSATD
 901 APRRRSRRSR RSLXLVTFPT SVESFRNLTG VNKLNGQGT FRFMSFLFGY
 951 RSDKLKLAES SEGTYTLAVN NTCNEPASLE QLTVVEGKDN KPLSENINPT
 1001 LQNEHVDAGA W.....
 //
 1151LDRVFAEDR
 1201 RNAWTSRQDI DTNRYNGQDF RAYGQDTDLR QIGGQKHLGS GRVGLFESHN
 1251 RTEMTFDDGI GNSARLAFGA VFGGVDIDRF YIGISAGAGF SSGSLSDIIG
 1301 KKKRRRVLHY GIQARYRAGE GGGFIEPHIG ATRYVQKAD YRYENVNIAI
 1351 PGLAFNRVYA GIKADYSEKP AQHISITPYL SLSYTDAASG KVRTRVNTAV

This corresponds to the amino acid sequence <SEQ ID 648; ORF1>:

1401 LAQDFGKTRS AEWGVNAEIK GFYLSLHAAA AKGPFQLEAOH SAGIKLYRW
1451 *

Further sequencing analysis revealed the complete nucleotide sequence <SEQ ID 649>:

5 1 ATGAARACAA CCGACAAACG GACAACGGAA ACACACCGCA AAGCCCGGAA
51 AACCGCGCGC ATCCCGCTTCT CGCGTGCTTA CTTCAGCATTA TGCCTGTGCT
101 TCGGCATTCT TCCCAACGCG TGGCGGGGAC ACACCTATTT CGGCATCAAC
151 TACCAATACT ATCGCGSACT TGGCCAAAAT AAGGCAAGT TTGCACTCGG
201 GCGCAAGAGT ATTGAGGTTT ACACAAAAAA AGGGGAGTTG GTCCGCAAAAT
10 251 CAATGACAAA AGCCCGCATG ATTGATTTTT CTGTGGTGTG CGGTACACCG
301 GTGGCGGCAT TGTGGGCGCA TCAATATATT GTGAGCGTGG CACATAACGG
351 CGGCTATAAC AACGTTGATT TTGGTGCGGA AGGAAGAAAT CCCGATCAAC
401 ATCGTTTTAC TTATAAAATT GTGAAACGGA ATAATTATTA AGCAGGGAAT
451 AAGGGCCATC CTTATGCGCG CGATTATCAT ATCGCGGTT TGCATAAATT
501 TGTACAGATG GCAGAACCTG TTGAATGAC CAGTTATATG GATGGGCGGA
15 551 AATATATCGA TCAARAATAAT TACCGTGACC GTGTTGCTAT TGGGCGAGCG
601 AGGCAATATT GCGCATCTGA TGAAGATGAG CCCAATAACC GCGAAAGTTT
651 ATATCATATT GCAAGTGGCT ATCTTTGGCT CGTGTGGTGG AATACCTTTG
701 CACAAAATGG ATCAGTGGT GGCACAGTCA ACTTAGTAG TGAAAAAATT
751 AAACATAGCC CATATGGTTT TTTCACACCA GGAGGCTCAT TTGGCAGACG
20 801 TGCGTACCCA ATGTTTATCT ATGAGGCCA AAGCAAAAG TGGTAAATTA
851 ATGGGGTATT GCAAAACGCG AACCCCTATA TAGGAAAAG CAAATGGCTT
901 CAGCTGGTTC GTAAGAGTTG GTTCTATGAT GAATCTTTG CTGAGATAC
951 CCATTCACTA TTCTACGAAC CAOGTCAAAA TGGGAARTAC TCTTTTACG
25 1001 ACGATAATAA TGGCAGCGGA AAAATCAATG CCAACATGA ACACATTTCT
1051 CTGCTAATA GATTAARAAAC ACGAACGTTT CAATTGTTTA ATGTTTCTTT
1101 ATCCGAGACA CAAAGAGAAC CTGTTATACA TGCTCGAGGT GGTGTCAACA
1151 GTTATCGACC CAGACTGAAT AATGGAGAAA ATATTTCTTT TATTGACGAA
1201 GGAAGAGGCG AATTGATACT TACCAAGCAAC ATCAATCAAG GTGCTGGAGG
1251 ATTATATTTC CAAGGAGATT TTACGGTCTC GCGTGAATAT AACGAACTT
1301 GCGAAGGCGC GGGGTTTCAT ATCAGTGAAG ACAGTACCGT TACTTTGAAA
30 1351 GTAAACGGCG TGGCAAAACG CCGGCTGTCC AAAATCGGCA AAGGCAGCT
1401 GCACGTTCAA GCCAAGGGGG AAAACCAAGG CTGATCAGC GTGGGCGAGC
1451 TACAGCTCAT TTTGATGACG CAGGCAAGCG ATAAAGGCAA AAAACAGGCC
1501 TTTAGTGAAA TCGGCTTGGT CAGCGCGGAG GTGACGCTGC AACTGAATGC
35 1551 CCAATATCAG CAAATGCTAC ACAACTCTTA TTTCGGCTTT CGCGCGGAC
1601 GTTGGAGATT AAGAGGCGAT TCGGTTTCTG TCAACCGTAT TCAAAATACC
1651 GATGAAGGGG CGATGATTGT CAACCAATAT CAAGCAAAAG AATCCACGCT
1701 TACCAATTACA GGCATAAAG ATATTGCTAC AACCGCAAT AACACAGCT
40 1751 TGGATAGCAA AAAAGAAATT GCTCAACAG GTTGGTTGG CAGAAAGAT
1801 ACGACCAAAA CGAACGGGCG GCTCAACCTT GTTTCACAG CCGCCGAGCA
1851 AGACCGCACC CTGCTGCTTT CCGCGGGAAC AATTTTAAAC GGCAACATCA
1901 CGCAACAAAA CGGCAACACT TTTTTCAGCG CGACCAACAC ACGCAGCGCC
1951 TACAATCATT TAAACGACCA TTGGTGCGR AAGAGGGGCA TTCTCTCGGG
2001 GGAATTCGTG TGGGACACAG ACTGGATCAA CGCACATTTT AAAGCGGAAA
45 2051 ACTTCCAAAT TAAAGGCGGA CAGCGGCTGG TTTCCGCGAA GTTGGCAAA
2101 GTGAAGAGCG ATTGGCATTT GAGCAATCAC GCCAACGAG TTTTGGTGT
2151 CGCACGCGAT CAAAGCCACA CAATCTGTAC ACGTTCGAG TGAACGGGTC
2201 TGACAARIITG TGTGAAAAAA ACCATTACCG ACGATAAAGT GATTGCTTCA
50 2251 TTGACTAAGA CGGACATCAG CGGCAATGTC GATCTTGGC ATCAGCGTCA
2301 TTTAAATCTC ACGGGGCTTG CCACACTCAA CGGCAATCTT AGTGCAATGT
2351 GCGATACACG TTATACGCTT CCGACAGAG CACCCAAAA CGGCAACCTT
2401 AGCGCTGTCG CGATGCCCA AGCAACTTTT ATTCACGCA CATTAAACCG
2451 CACACATCTG CTTTGGGCGA ATGCTTCATT TAATCTTACT GACCAACGCG
55 2501 TACAAAAACG CAGTCTGAAG CTTCGCGCA ACGCTAAGG AAGCTTAGC
2551 CATTCGCGAC TCAACGGTAA TGTCTCCCTA GCGATTAAGC CAGTATTCCA
2601 TTTGAAAGC AGCGCTTTTA CCGGACAAAT CAGCGGCGCG AAGGATACGG
2651 CATTACACTT AARAGACAGC GAATGAGCGC TGCCTCAGG CAGGAAATTA
2701 GGCAATTTAA ACCTTGACAA CGCCACCACT ACACCTAATT CCGCTATCG
2751 CACGATGCG CAGGGGCGCG AAACCGGCG TCGACAGAT GCGCGCGCGC
60 2801 GCGGTTGCGC CGGTTGCGCG CGTTCCTTAT TATCGGTTAC ACGCGCAACT
2851 TCGTGTAGAT CCGGTTTCAA CAGCTGAGC GTAAACGGCA AATTGAACGG
2901 TCAGGGAACT TTTCCGTTTA TGTGGAAGCT CTTGCGTAC CGCAGCGACA
2951 AATTGAAGCT GCGGGAAGAT TCGGAAGGCA CTTACACTTT GCGGTCACAC
3001 AATACCGGCA ACGAACCTGC AAGGCTCGAA CAATTGAGCG TAGTGAAGG
65 3051 AARAGACAAAC AAACCGCTGT CCGAAAACCT TAATTTCAAC TGCARAAACG
3101 AACACGTCGA TGCGGGCGCG TGGGTTTACC AACTCATCGC CAAGACCGCG
3151 GAGTTCCGCC TGCAATATCC GTCTCAAGAA CAGAGACTTT CGCAACACT
3201 CCGGACAGCA GACGCAAAA AACAGGCGGA AAGAGCAAC GCGCAAGCC
3251 TTGACGCGCT GATTGCGGCG GCGCGGAGT CCGTCAAAA GACACAAAGC
70 3301 GTTGCAGAAC CGCCCGCGCA GGCAGGCGGG GAAATATGCG GCATTATGCA

3351	GGCGGGAGAA	GAGAAAAAAC	GGGTGCAGGC	GGATAAAGAC	ACCGCCTTGG
3401	CGAAGACGCG	CGAGACGGA	ACCGGCGCG	CTTCCACCGC	CTTCCCGCGC
3451	GGCCGCCGCG	CGCCCGGGA	TTTGCCTGCA	CTGCAACCCC	AACCGCAGCC
3501	CCAACCGCAG	CGCGACTGA	TCAGCCGTTA	TGCGAATAGC	GGTTTAGATG
3551	AATITCCGCG	CAGCCTCAAC	AGCGTTTTCG	CGGTACAGGA	CGAATTAGAC
3601	CGCCTATTTC	CGGAAGACCG	CGCGAACGCC	GTTTGGACAA	CGCGCATTCG
3651	GGACACCAAA	CACCTACGGT	CGCAAGATTT	CCGCGCCTAC	CGCCACAATA
3701	CGGACCTGGC	CGAATTCGGT	ATGCAGAAAA	ACCTCGGCAG	CGGCGCGGTC
3751	GGCATCTGT	TTTCGCACAA	CCGAGCCGAA	AACACCTTCG	ACGACGGCAT
3801	CGGCAACTCG	CGACGGCTTG	CCACGGCGCG	CGTTTTCGGG	CAATACGGCA
3851	TCGACAGGTT	CTACATCGGC	ATCAGCGCGG	GGCGGGGTTT	TAGCAGCGGC
3901	AGCCTTTCAG	ACGCGACTCG	AGCGAAATTC	CGCGCGCGCG	TGCTGCATTA
3951	CGGCATTTCAG	CGACGATACC	CGCGCGGTTT	CGCGGGATTC	GGCATCGAAC
4001	CGCACATCGG	CGCAACGCGC	TATTTCTGCC	AAAAAGCGGA	TTACCGCTAC
4051	GAJAAAGTCA	ATATCGCGAC	CCCCGGCCTT	GCATTCAACC	GCTACCGCGC
4101	GGGCATTAA	CGAGATTATT	CATTCAAAAC	GGCGCAACAC	ATTTCCATCA
4151	CGCCTATTAT	GAGCCTGTTC	TATCAGGATG	CGCGCTCGGG	CAAGCTCCGA
4201	ACACGGCTCA	ATACGGCGGT	ATTGCGCTAG	GATTTCCGCA	AACCCGCGAG
4251	TGCGGAATGG	GGCTAAACG	CGGAATCAA	AGGTTTCCAG	CTGTCCTCC
4301	ACGCTCGCGC	CGCCAAAGGC	CCGCAACTGG	AAGCGCAACA	CAGCGCGGGC
4351	ATCAAAATTAG	GCTACCGCTG	GTA		

This corresponds to the amino acid sequence <SEQ ID 650; ORF1-1>:

1	MKTTDKRTTE	THRKAPKTGR	IRFSPAYLAI	CLSFGLPOA	WAGHTYFGIN
51	YQYYRDFAE	KGKFAVGA	KDIEVYNK	KGELVGS	MTKAPMID
101	VAALVGDQYI	VSAHNGGYN	NVDFGAEG	RNPDQHRF	FTYKIKR
151	KGHPYGGDYH	MPRLHKFVF	AEPEVMSY	MDGRKIDQ	NPYDRVIR
201	RQYWRSDDE	PNNRESSYH	ASAYSWL	VGGNTFAQ	NGSGGTV
251	KHSPYGLFPT	GGSGDGS	SPMFIYDA	QKQKWL	INGVLQ
301	QLVRKDWDF	EIPAGDTH	SVFYEP	PRONGK	YSFNDDN
351	LPNRLKTRTV	QLENVSL	SETARE	PEVYHA	AGVNSY
401	KGKGLILTSN	INQAGGLY	FQGDFT	VSPEN	NETWQAG
451	VNGVANDRLS	KIGKGLH	LVQAKEN	QCGSIS	VGDGTV
501	FSEIGLVSGR	GTVLNAD	NGFNPDK	LYFG	RGRGLD
551	DEGAMIVNHN	QKESTVT	ITGNKDKI	ATTGN	NSNLS
601	TTKTRGRLNL	VQPAEDRT	LLLSGGT	NLW	GNITGTY
651	YNHLNDHWSQ	KEGIPRGE	IVDNDW	INRTE	KAENFQ
701	VKGDWHLNSH	AQAVFGV	APHOSHT	ICTRSD	WTGLTNC
751	LTKTDDSGNV	DLADHAHL	NLTLGLAT	LNGL	SANGDTRY
801	SLVNGAQT	FNRQAT	LNNGNTS	ASGNASE	NLS
851	HSALNGVSL	ADKAVF	HESRFT	GQISGG	KDTALH
901	GNLNLNATI	TLNSAYR	HDAAQAT	CGSATD	APRRRS
951	SVESRENTLT	VNGKLN	CGQTF	FRFMS	ELFGY
1001	NTGNEPASLE	QLTVVEG	KDNKPL	SENLFN	TQLNEH
1051	EFLRLNPVKE	QELSDKL	KCAEAKQ	AEKDN	AQSLDAL
1101	VAEPAQAGG	ENVGIMQ	AEE	EKKRVQ	ADKD
1151	ARRARDLPQ	LQPOPO	POPO	RDLISRY	ANS
1201	RVFAEDRRNA	VMTSG	IRDTK	HYRSQD	FRAY
1251	GILFSHNRTK	NTFDDG	IGNS	HLA	HAVFG
1301	SLSDGIGGKI	RRRVLY	HIGQ	ARYRAG	FGFG
1351	ENVNIATPGL	ENFRYR	AGIK	ADYSEK	PAQH
1401	TRVNTVLA	Q	DFGKTR	SAEW	GVNAEIK
1451	IKLGYRW*				

Computer analysis of these sequences gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

- 55 ORF1 shows 57.8% identity over a 1456aa overlap with an ORF (ORF1a) from strain A of *N. meningitidis*:

	10	20	30	40	50	60
orf1.pep	MKTTDKRTTET	THRKAPKTGR	IRFSAAYLA	ICLSFGLPQ	AWAGHTYFG	INYYRDFAE
60 orf1a	MKTTDKRTTET	THRKAPKTGR	IRFSAAYLA	ICLSFGLPQ	AWAGHTYFG	INYYRDFAE
	10	20	30	40	50	60
	70	80	90	100	110	120
orf1.pep	KGKFAVGA	KDIEVYNK	KGELVGS	MTKAPMID	FSVSVSRNG	VAALVGVQYI

-363-

```

|||||
orfla  GDTRYTVSHNATQNGNLSLVGNAQATFNQATLGNKXSGNASFNLSNNAAGNSLTLSLD
      780      790      800      810      820      830

5
      550      560      570      580      590      600
orfl.pep NAKANVSHSALNGNVSLADKAVHFESSRFTGQISGGKDTALHLKDESWTLPSGXELGNL
      |||||
orfla    NAKANVSHSALNGNVSLADKAVHFESSRFTGQISGGKDTALHLKDESWTLPSGTLELGNL
      840      850      860      870      880      890

10
      610      620      630      640      650      660
orfl.pep NLDNATITLNSAYRHDAAGAQTGSATDAPRRRSRRRSLLXVTPPTSVESRFNTLTVNG
      |||||
orfla    NLDNATITLNSAYRHDAAGAQTGVSDTFRRRSRRS---LLSVTPPTSVESRFNTLTVNG
      900      910      920      930      940      950

15
      670      680      690      700      710      720
orfl.pep KLNKGQGTFRFMSELFGYRSDKLKLAESESGTYTLAVNNTGNEPVSLEQLTVVEGKDNKPL
      |||||
orfla    KLNKGQGTFRFMSELFGYRSDKLKLAESESGTYTLAVNNTGNEPVSLEQLTVVEGKDNKPL
      960      970      980      990      1000      1010

20
      730      740      750
orfl.pep SENLNFTLQNEHVDAAGW-----
      |||||
orfla    SENLNFTLQNEHVDAAGWRYQLIRKDGFEFLHN PVKEQELSDKLGKAEAKKQAEKDNAS
      1020      1030      1040      1050      1060      1070

25
orfl.pep -----
orfla    -----

30
orfl.pep -----
orfla    LDALIAAGRDAAEKTESVAEPARXAGGENVGIMQAEEEKKRVQADKDSALAKQREAEATRP
      1080      1090      1100      1110      1120      1130

35
orfl.pep -----760
orfla    XTTFAPFRARXARDLPQOPQOPQOPQOPQORDLXSRYANSGLSEFSATLNSVFAVQDELDR
      1140      1150      1160      1170      1180      1190

40
      770      780      790      800      810      820
orfl.pep VFAEDRRNAVVTSGIRDTKHYRSQDFRAYRQOTDLRQIGMQKNLGSGRVGILFSHNRTE
      |||||
orfla    VFAEDRRNAVVTSGIRXTKHYRSQDFRAYRQOTDLRQIGMQKNLGSGRVGILFSHNRTE
      1200      1210      1220      1230      1240      1250

45
      830      840      850      860      870      880
orfl.pep TFDDGIGNSARLAHGAVFQGYGIDRFYIGISAGAGFSSGSLSDGIGKXKRRVLHYGIQA
      |||||
orfla    XFDDGIGNSARLAHGAVFQGYGIGRFDIGISTGAGFSSGSLSDGIGKIRRVVLHYGIQA
      1260      1270      1280      1290      1300      1310

50
      890      900      910      920      930      940
orfl.pep RYRAGFGGFGIEPHIGATRYFVKADYRYENVNIATPGLAFNRYRAGIKADYSFKPAQHI
      |||||
orfla    RYRAGFGGFGIEPYIGATRYFVKADYRYENVNIATPGLAFNRYRAGIKADYSFKPAQHX
      1320      1330      1340      1350      1360      1370

55
      950      960      970      980      990      1000
orfl.pep SITPYLSLSTDAASGKVRTRVNTAVLAQDFGKTRSAEWGNVNAEIKGFTLSLHAAAAAGKP
      |||||
orfla    SITPYLSLSTDAASGKVRTRVNTAVLAQDFGKTRSAEWGNVNAEIKGFTLSLHAAAAAGKP
      1380      1390      1400      1410      1420      1430

60
      1010      1020
orfl.pep QLEAQHSAGIKLGYRWX
      |||||
orfla    QLEAQHSAGIKLGYRWX
      1440      1450

```

70 The complete length ORF1a nucleotide sequence <SEQ ID 651> is:

1 ATGAAACAA CCGACAAAG GACACCGAA ACACACGGCA AAGCCCGGAA
 51 AACCGCGCG ATCCGCTTA CGCTGCTTA CTTCAGCAT TGCCTGCTGT
 101 TCGCATCTCT TCCCGAGCT TCGCGGGAG ACACCTATT TCGGCATCAAC
 151 TACCARTACT ATCGCGACT TCGCGAAAT AAGGCAAGT TTGCACTCGG
 201 GCGCAAGAT ATTGAGTNT ACAACAAAA AGGGGAGTTG GTGCGCAAT
 251 CAATGACAAA AGCCCGCATG ATTGATTTT CTGTGGTTC GCGTAACGCG
 301 GTGGCGGCAT TGGTGGGCGA TCAATATATT GTGACGCTGG CACATAACGG
 351 CGGCTATAAC AAGCTTGATT TTGGTGGCGA AGGAAGNAAT CCGCATCAGC
 401 ACCGTTTTTC TTACCAAAAT GTGAAAGAA ATAATTATAA CCTGACAAAT
 451 TCACACCCCT ACACGCGCGA TTANCATATG CCGCGTTTGC ATAAATTTGT
 501 CACAGATGCA GAACCTGTGC AAATGACGAG TGACATGAGG GGAATATACCT
 551 ATTCOGATAA AGAAAATAT CCGAGCGTG TCGCATCGG CTCAGGACAC
 601 CACTATTGGC GTTATGATGA TGACAAACAC GCGCATTTAT CCTACTCCGG
 651 GCGATGGTTA ATTGGCGGCA ATACACATAT GCAGGTTTGG GGAATAAATG
 701 GCGTANTTAG TTGAGCGGCG GATGTGCGCC ATGCCAACGA CTATGGCCCT
 751 ATGCGGATTG CAGGTGCGGC AGGCGACAGC GGTTCGCCAA TGTTTATTTA
 801 TGACAAACAA AACATATAAT GGTGCTCAA CGSAGTTTTA CAACCGGGCT
 851 ACCCTTATTG CGCAGCGAA AAGCTTTTCC AGCTGTATG CAAGGTTGCG
 901 TTCTACGATG ACATTTACAG AGGCGATACA CATACGCTCT NTTTTGAACC
 951 GCGCAGTAAC GGACATTTTT CTTTACATC CACACAAAC GTTACGGGTA
 1001 CGGTAAACGA AACCAACGAA AAGGNTCCA ATCCAAAGCT TAAAGTACAG
 1051 ACAGTCCGAC TGGTTGACGA ATCTTTGAAT GAACTGATA AAGAACCAGT
 1101 TTACGCGGCA GGGGTTGTTA ATCAGTACCG TCCAAGTTA AACAAACGGT
 1151 AAAACCTTTC TTTTATCGAT TACGGCAACG GCAAACTCAT CTATCAAAAC
 1201 AACATCAACC AAGCGCGGG CGGTTTGTAT TTTGAAGGTG ATTTTACGGT
 1251 CTCGCTGAA AACACGAGAA CGTGGCAAGG CCGCGCGGTT CATATCAGTG
 1301 AAGACAGTAC CGTTACTTGG AAGTAACGCG GCGTGGCMAA CGACCGCTTG
 1351 TCCAAATCG CCAAAAGGAC GCTGCAAGT CAAGCCAAAG GGGAAACCA
 1401 AGGCTCGATG AGCGTGGGCG ACGGTACAGT CATTTTGGAT CAGCAGGACG
 1451 ACGATAAAAGG CAAAAACAA GCCTTTAGTG AAATCGGCTT GNTCAGCGCG
 1501 AGGGGTACGG TGCAACTGAA TGGCGATAAT CAGTTCAAAC CCGACAAACT
 1551 CTATTTCCGC TTTGCGGGCG GAGCTTTGGA TTTAAACGGG CATTCGCTTT
 1601 CTTTCCACCG TATTCAAAAT ACGGATGAAG GGGGATGAT TGNMTCATCT
 1651 AATCGCACAA AACATCAAC CTTTACATG ACAGGAAAG AAGTATTAC
 1701 ACACCGAGT GGTGAAGATA TCAATAGACT TANTTACAGC AAGAATATG
 1751 CCTACAACGG TTGTTTGGC GAGAAGATA CAGCAAAAC GACAGGGCG
 1801 CTCACCTTGG TTTACAGGCC CGCCGAGAA GACCGCACCC NGCTGCTTTC
 1851 CGCGGGAACA AATTTAAAGC GCAACATCAC GCAAAACAAAC GGCAACTGT
 1901 TTTTACGCGG CAGACCGACA CCGCACGCTC ACAATCATTT AGGAACGCGG
 1951 TGGTCAAAAA TGGAAGGTAT CCGACAAGGA GAAATCGTGT GGGACAACGA
 2001 CTGGATCNAC CGCAGCTTTA AAGCGGAAAA TTTCCATATT CAGGGCGGGC
 2051 AGGCGGTGAT TTCCGCAATG GTTGCCAAAG TGGAAGCGGA TTGNCATTGT
 2101 AGCAATCAGC CCAAGCAGT TTTTGTGTCT GCACCGCATC AAAGCCATAC
 2151 AATCTGTACA CGTTGGGACT GGACNGTCT GACAAATTTG GTCGAANAAA
 2201 NCATTACCGA CGATAAAGTG ATTGCTTCAT TGACTAAGC NGACNTNAGC
 2251 GGCANTGTNA GNTCNMCCNA TNACGNTNNT NTAACNTCN CNGGGCTGCG
 2301 NNCACNTAAN GGCAATCTTA GTGCAAAATG CGATACAGT TATACAGCTCA
 2351 GCCACAACGC CACCCAAACG GGCAACCTTA GCTCGTGGG CAATGGCCAA
 2401 GCACATATTA ATCAAGCGAC ATTAACCGGC AACNCATCGG NTTGCGGAAA
 2451 TGCTCTATTG AATCTAAGCA ACAGCGCCCG ACACAAACGCG AGTCTGACGC
 2501 TTTCGAGACA AAGCTAAGCG ATTCGAGCT CAAGCGAAT
 2551 GTCTCCCTAG CCGATAAGGC AGTATTCCAT TTGCAAAAAC GCGCCTTAC
 2601 CGGCACTACT AGCGCGCAGCA AGGANACAGC ATTACATTA AAGACAGCG
 2651 AATGAGCGCT CGCGTCAGGC ACGGAATTAG GCAATTTAAA CTTTGACAC
 2701 GCCACATTAA CACTCAATTC CGCCTATCGC CACGATGCTG CAGCGCGCGA
 2751 AACCGGAGN GTGTGACACA CGCCGCGCGC CGGTTGCGGC CGTTCCTAT
 2801 TATCGGTTAC ACGCGCAACT TCGGTAGATAT CCGGTTTCAA CACGCTGACG
 2851 GTAAACGGCA AATTGAACNG TCAAGGAACA TTCGCTTTA TGTGGAACAT
 2901 CTTGCGGTAC CGAAGCGACA AATTGAAGCT GCGCGAAGT TCCGAAGGNA
 2951 CTTACACCTT GCGCGTCAAC AATACCGGCA ACGAACCCTG AAGCTCGAT
 3001 CAATTGACGG TAGTGGAAAG GAAAGACAAC AACCGCTGT CCGAAAACCT
 3051 TAATTTACCC CTGCAAAAGC AACACGTCGA TCGCGGGCGT TGGCGTTACC
 3101 AACTCATCGC CAAGAACGCG GAGTTCGCGC TGCAATATCC GGTCAAGAA
 3151 CAAGAGCTTT CCGACAAACT CCGCAAGGCA GAAGCCAAAA AACAGCGGGA
 3201 AAAAGACAA CCGCAAGGCC TTGACGCGCT GATTGCGGCC GGGCGGATG
 3251 CCGCGGAAAA GACGAAAGC GTTGGCGAG CCGCGCGGCG GGCAGGCGG
 3301 GAARATGTCG GCATTTGCGA GCGCGAGGAA GCGCGAGGAA GAGAAAGGAA
 3351 GGAATAGAC AGCGNMTTG GCGCGAGGAA GCGCGAGGAA ACCGCGGCG
 3401 NTACCACGCG CTTCCCGCGC GCGCCGCGCG CCGCGGGGA TTTGCGGAA
 3451 CCGCAGCCCC AACCGCAACC TCAACCCCAA CCGAGCGCGC ACCTGATNAG
 3501 CCGTTATGCC AATAGCGGTT TGAGTGAATT TTCGCGCAG CTCAACAGCG
 3551 TTTTCGCGGT ACAGGACGAA TTGACCGCG TGTGTGCGGA AGACCGCCCG

5
3601 AACGCGNTTT GGACAAGCNG CATCCGGNAC ACCAAACACT ACCGTTCCGA
3651 AGATTTCACG CTCTACACGC CTTGCGCCAA ATCGGTATCGC
3701 AGAAAAACCT GCGACCGGG CGGTCGGCA TCGCTTTTC GCACACCGGG
3751 ACGAAACA NCTTCGACGA CGGCATCGCG AACTCGGCAG GCGTTGCCA
3801 CGCGCCGCTT TTCCGGCAAT ACGCCATCGG CAGGTTGAC ATCGGCATCA
3851 GCACGGGGCG GGGTTTTCAG AGCGGCANTC TNTCAGACCG CATCGGAGGC
3901 AAAATCCGCC GCGCGTGTCT GCATTACCGG ATTCAGGACG GATACGCGCG
3951 CGGTTTCGGC GGATTTCGCA TCGAACCGTA CATCGGCGCA ACGCCTATT
10
4001 TCGTCCAAAA AGCGGATTAC CGCTACGAAA ACCTCAATAT CGCCACCCCC
4051 GGTCTTCGCT TCAACCGNTA CGNCGGGGCG ATTAAGGCGC ATTATTCATT
4101 CAAACCGCGC CAACACATNT CCATCAACNC TTATTTNAGC CTGTCCTATA
4151 CGGATCGCGC TTCCGGCAAA GTCGGAACAC GCGTCAATAC CGCNGTATTG
4201 GCTCAGGATT TCGGCAAAAC CCGCAGTTCG GAATGGGGCG TAAACGCCGA
4251 AATCAAAGGT TTCACTGTGT CINTCCAGCG TCGCGCGCGC AAAGNCCGC
15
4301 AACTGGAAGC GCACAACAGC GCGGCGATCA AATTAGGCTA CGGCTGGTAA

This encodes a protein having amino acid sequence <SEQ ID 652>:

1 MKTTDKRRTTE THRKAAPTGR IRFSPAYLAI CLSFGILPOA WAGHTYFGIN
51 YQYRDFPAEN KGKFAVGAKD IEVYNKKGEL VGKSMTKAFM IDFSVVSRRNG
101 VAALVGQDYI VSVAHNGSYN NVDFGAEGXN PDQHRFSYQI VKRNNYKFPN
151 SHPVGDGDM PRLLHKFYTDK EPVWETSDMR GNTYSDKKEI PPRVIGSGH
201 HWRYDDDKH GDLSYSGAWL IGGNTHMQGW GNNGVXSLSG DVRHANDYGP
251 MPIAGAAGDS GSPMFIYDKT NKNKLLNGVL QGTGYPYSGRE NGFQLIRKDW
301 FYDDIYRGDT HTVXFEPFSN GHFSFTSNNN GTGTVTETNE KVSNNPKLKVO
351 TVRLFDSESLN ETDKEPYVAA GGVOYRFRLL NNGENLSFTD YGNKGLILSN
401 NINQAGAGLY FEGDFTVSP E NNETWQAGAV HISEDSTVTV KVNQVANDRL
451 SKIGKGTLLHV QAKGNGQSGI SVGDGTVILD QQADGKGKQ AFSEIGLXSG
501 RGTVQLNADN QFNPDKLYFG FRGRLDLNG HSLSFHRIQ TDEGAMIXXH
551 NATTTSTVTI TGNEISITQPS GKINIRLNYS KEIAYNGWFG EKDDTKTNKR
601 LNLVYQPAEE DRTXLLSGGT NLNGNITQTN GKLFYSGRPT PHAYNHLGSG
651 WSKMEGIPQG EIVWENDWIX RTFKAENFHI QGGQAVISRN VAKVEGDHXL
701 SNHAQAVFGV APHOSHITCT RSDWTGLTNC VEXXITDDKV IASLTXTDKS
751 GXVLLXXXXX XXLLGXAKLX GNLSANGDTR YTVSHNATQN GNLSLVGNAQ
801 ATPNQATLNG NXSXSGNASF NLSNNAQNG SLTSLDNAKA NVSHSALNG
851 VSLADRAVPH FENSRTFGQL SSGKXTALHL KDSMTLPFG TELGNLNLN
901 ATITLASAYR SDRAAQPCX VSDTERRSR RSLLSVFPPT SVESRNTLT
951 VNGKLNQGT FRFMSLEFG RSDKLKLAES SECTYTLAVN NTGNEPVS LD
1001 QLTVEGKDN KPLSENINFT LQNEHVDAGA WRYOLIRKDG ERLHNFVKE
1051 QELSOKLKA EAKKQAEKN AQSLDALIAA GRDAEKTES VAEPARXAGG
1101 ENVGIMQAE EKKRVQADKD SALAKQREAE TRPYTTFAPR ARXARALDLP
1151 POPOPOPOPO PORDLXRYA NSGLSEFSAT LNSVFAODE LDRVFAEDRR
1201 NAVWTSXIRX TKHYSQDFR AYRQOTDLRG IQMQRNLGSG RVGILFSHNR
1251 TENXFDGIG NSARLAHAGV FGQYIGRFD IGISTGAGFS SGXLSDGIGG
1301 KIRRRVLHYG IQARYRAGFG GFGIEPYIGA TRYFVQADY RYENVNIAPT
1351 GLAFNRYRAG IKADYSFKPA QHSITPYXS LSYTDAASGK VTRVNTAVL
45
1401 AQDFGKTRSA EWGVNAEIKG FTLSXHAANA KGFQLEAQS AGIKLYRW*

A transmembrane region is underlined.

ORF1-1 shows 86.3% identity over a 1462aa overlap with ORF1a:

		10	20	30	40	50	60
50	orf1a.pep	MKTTDKRRTTETHRKAAPTGRIRFSPAYLAICLSFGILPOAWAGHTYFGINQYRRDFPAEN					
	orf1-1	MKTTDKRRTTETHRKAAPTGRIRFSPAYLAICLSFGILPOAWAGHTYFGINQYRRDFPAEN					
		10	20	30	40	50	60
55	orf1a.pep	KGKFAVGAKDIEVYNKKGELVGKSMTKAFMIDFSVVSRRNGVAALVGQDYIVSVAHNGSYN					
	orf1-1	KGKFAVGAKDIEVYNKKGELVGKSMTKAFMIDFSVVSRRNGVAALVGQDYIVSVAHNGSYN					
		70	80	90	100	110	120
60	orf1a.pep	NVDFGAEGXNPDQHRFSYQIVKRNYYKPDNS-HPYNGDHMRPLRHKFTDAEPVEMTSDM					
	orf1-1	NVDFGAEGRNPDQHRFTYKIVKRNYYKAGTKGHYPGGDYHMRPLRHKFTDAEPVEMTSYM					
		130	140	150	160	170	179
65		130 140 150 160 170 180					

-366-

		180	190	200	210	220	230
	orf1a.pep	RGNTYSOKKEYPERVRI GSGHHYWRYYDDKNGDL--SYSGA---WLTGGNTHMQGSGNN					
5	orf1-1	DGRKYIDQNNYPDRVRI GAGROYRSDDEPNRRESSYHIASAYSWLVGGNTFAQNSGG					
		190	200	210	220	230	240
	orf1a.pep	240	250	260	270	280	290
10	orf1-1	GVXSLSGD-VRHANDYGPMP IAGAAGDSGSPMFIYDKTNNKWLINGVLQTGYPYSGRENG					
		250	260	270	280	290	
	orf1a.pep	300	310	320	330	340	350
15	orf1-1	FQLIRKDWFYDDIYRGDTHVXFEPSPRNGHFSFTSNNNGTGTVTETNEKVSNP-KLKVOT					
		300	310	320	330	340	350
	orf1a.pep	360	370	380	390	400	410
20	orf1-1	VRLFDESINETDKEPVY-AAGGVNQYRPLNNGENLSFIDYGNKGLISNNINQAGAGGLY					
		360	370	380	390	400	410
25	orf1a.pep	420	430	440	450	460	470
	orf1-1	FEGDFTVSPENNETWQAGVHISEDSTVTKVNGVANDRLSKIGKGTLHVQAKGENQSGSI					
30		420	430	440	450	460	470
	orf1a.pep	480	490	500	510	520	530
	orf1-1	SVGDGTVILDQQADDKGGKQAFSEIGLXSGRGTVQLNADNQFNPDKLYFGFRGRLDIMG					
35		480	490	500	510	520	530
	orf1a.pep	540	550	560	570	580	590
40	orf1-1	HSLSFHRIQNTDEGAMIXXHNATTTSTVTITGNESITQPSGKNINRLNYSKEIAYNGWFG					
		540	550	560	570	580	590
45	orf1a.pep	600	610	620	630	640	650
	orf1-1	EKDTTKTNGRLNLVYQPAEDRTXLLSGGTNLNGNITQTNGKLFSSGRPTPHAYNHLGSG					
		600	610	620	630	640	650
50	orf1a.pep	660	670	680	690	700	710
	orf1-1	WSKMEGIPQGEIWDNDWIXRTEKAENFHIQGGQAVISRNVAKVGGDKHLSNHAQAVFGV					
		660	670	680	690	700	710
55	orf1a.pep	720	730	740	750	760	770
	orf1-1	APHQSHITICTRSDWTGLTNCVEXXITDDKVIASLTKTDXSGXVLKXXXXXXLXGAXILX					
60		720	730	740	750	760	770
	orf1a.pep	780	790	800	810	820	830
65	orf1-1	GNLSANGDTRYTVSHNATQGNLSLVGNAQATFNQATLNGNIXKXSGNASFNLSNNAQNG					
		780	790	800	810	820	830
	orf1a.pep	840	850	860	870	880	890
70	orf1-1	SLTSLDNAKANVSHSALNGVSLADKAVFHFENSRTQTQLSGSKXTALHLKDESEITLPSG					
		840	850	860	870	880	890

		900	910	920	930	940
	orf1a.pep	TELGNLNLONATITLNSAYRHDAAGATGXVSOTPRRRSRSS---	LLSVTTPPTSVESRIN			
5	orf1-1	TELGNLNLONATITLNSAYRHDAAGATGXVSOTPRRRSRSSLLSVTTPPTSVESRIN				
		900	910	920	930	940
	orf1a.pep	950	960	970	980	990
		TLTVNGKLNQGTFRFMSELFYGRSDKLLAESSEGYTLAVNNTGNEPVSLDQLTVEG				
10	orf1-1	TLTVNGKLNQGTFRFMSELFYGRSDKLLAESSEGYTLAVNNTGNEPVSLDQLTVEG				
		960	970	980	990	1000
	orf1a.pep	1010	1020	1030	1040	1050
		KDNKPLSENINFTLQNEHVDAGAWRYQLIRKDGFRHLNPFVQEQLSKLGAEAARKQAE				
15	orf1-1	KDNKPLSENINFTLQNEHVDAGAWRYQLIRKDGFRHLNPFVQEQLSKLGAEAARKQAE				
		1020	1030	1040	1050	1060
	orf1a.pep	1070	1080	1090	1100	1110
		KDNAQSLDALIAAGRDAAEKTESVAEPARKAGGENVGIMQAEEEKKRVQADKDALAKOR				
20	orf1-1	KDNAQSLDALIAAGRDAAEKTESVAEPARKAGGENVGIMQAEEEKKRVQADKDALAKOR				
		1080	1090	1100	1110	1120
	orf1a.pep	1130	1140	1150	1160	1170
		EAETRPXTTAFPRARXARDLPQPQPQPQPQPORDLXSRYANSGLSEFSATINSVFVAV				
25	orf1-1	EAETRPXTTAFPRARXARDLPQPQPQPQPQPORDLXSRYANSGLSEFSATINSVFVAV				
		1140	1150	1160	1170	1180
	orf1a.pep	1190	1200	1210	1220	1230
		QDELDRVFAEDRRNAVWTSXIRKTKHYRSODFRAYRQQTDLRQIGMQNGLSGRGVILFS				
30	orf1-1	QDELDRVFAEDRRNAVWTSXIRKTKHYRSODFRAYRQQTDLRQIGMQNGLSGRGVILFS				
		1200	1210	1220	1230	1240
	orf1a.pep	1250	1260	1270	1280	1290
		HNRTENXFDGIGNSARLAHGAVFQYQIGRFDIGISTGAGFSSGLSDGIGKIRRRVL				
35	orf1-1	HNRTENXFDGIGNSARLAHGAVFQYQIGRFDIGISTGAGFSSGLSDGIGKIRRRVL				
		1260	1270	1280	1290	1300
	orf1a.pep	1310	1320	1330	1340	1350
		HYGIQARYRAGFGGFGIEPIGATRYFVQKADRYENVINIATPGLAFNRYRAGIKADYSF				
40	orf1-1	HYGIQARYRAGFGGFGIEPIGATRYFVQKADRYENVINIATPGLAFNRYRAGIKADYSF				
		1320	1330	1340	1350	1360
	orf1a.pep	1370	1380	1390	1400	1410
		KPAQHXSITPYLSYSYTDAAASKVTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHA				
45	orf1-1	KPAQHXSITPYLSYSYTDAAASKVTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHA				
		1380	1390	1400	1410	1420
	orf1a.pep	1430	1440	1450		
		AAAKGPQLEAQHSAGIKLGYRWX				
50	orf1-1	AAAKGPQLEAQHSAGIKLGYRWX				
		1440	1450			

Homology with adhesion and penetration protein hap precursor of *H. influenzae* (accession number P45387)

Amino acids 23-423 of ORF1 show 59% aa identity with hap protein in 450aa overlap:

orf1	23	FXAAYLAICLSFGILPQAWAGHTYFGINYYQYRDAENKGGKFAVGAKDIEVYNKKGELVG	82
		F +L C+S GI QAWAGHTYFGI+YQYRDAENKGGK VAK+IEVYNK+G+LWG	
65	hap	FRNLFLTACVSLGIAQAWAGHTYFGIDYQYRDAENKGGKTVGAKNIEVYNKGGQLVG	65
orf1	83	KSMTPKAMIDFSVVSRRNGVAALVGQYIVSVVHNGGYNVDFGAECNIXDQKRTYKIV	142
		SMTKAMIDFSVVSRRNGVAALVG QYIVSVVHNGGYN+VDFGAEG N DQ R TY+IV	
hap	66	TSMTKAMIDFSVVSRRNGVAALVGQYIVSVVHNGGYNVDFGAEGRN-PDQHRFTYQIV	124

5	orf1	143	KRNYYKAGTKGHPYGGDYHMPRLHKXVTDAEPVEMTSYMDGRKYIDQNNYDPVRIGAGR	202
	hap	125	KRNYY+A + HPY GDYHMPRLHK VT+AEVP MT+ MDG+ Y D+ NYP+VRIG+GR	184
10	orf1	203	QWRSDEDEPNNRESSYHIA-----	222
	hap	185	QYWR+D+DE N SSY+++	244
15	orf1	278	--YEPRNGKYSFNDNNGTGKIN-AKHEHNSLPNRLKTRTVQLFVNSLSETAREPVYHA	334
	hap	305	Y P NG YSF +N+GTGK+ + + + TV+LFN SL++TA+E V A	363
20	orf1	335	AGGVNSYRPLNNGENISFIDEGKGLILTSNINOGAGGLYFGQDFTV-SPENNETWQGA	393
	hap	364	A G N Y+PR+ G+NI D+GK L + +NINOGAGGLYF+G+F V +NN TWQGA	423
	orf1	394	GVHISEDSTVTWVNGVANDRLSKIGGTL 423	
	hap	424	GV I +D+TV WKV+ NDRLSKIG GTL	453
25	Amino acids 715-1011 of ORF1 show 50% aa identity with hap protein in 258aa overlap:			
30	Orf1	41	DTRYVTSNATQ-NGNXSLVXNAQATFNO-ATLNGNTSASGNASFNLSDHAVQNGSLTSL	98
	hap	733	DT+ S TQ NG+ +L NA + A LNGN + ++ F LS++A Q G++ LS	792
35	orf1	99	GNAKANVSHSALNGVNSLADKAVHFESSRFTGQISGGKDTALHLKDESWTLPSGXELGN	158
	hap	793	+A A V+++ LNGVN L D A F ++S F QI G KDT + L+++ WT+PS L N	852
40	orf1	159	LNLDAITILNSAYRHDAAQAQGTSGATDAPKXXXXXXXXXXLLKVTPTPTSVERNLTIVN	218
	hap	853	L L N+T+TLNSAY + S +AP L T PTS E REN+LTIVN	899
45	orf1	219	GKLNQGTFRFMSELFGYRSDKLLAESSEGTYYLAVNNTGNEPASLEQLTVVEGKDNKP	278
	hap	900	GKLGQGTFF+P S LFGY+SDKLLK+ +EG YTL+V NTG EP +LEQLT++E DNKP	959
50	orf1	279	LSNINFTLQNEHVDAGA 296	
	hap	960	LS+ L FTL+N+HVDAGA	977
55	orf1	1	LDRVFAEDRRNAVVTSGIRDTHKHYRSQDFRAYRQQTDLRIGMQKNLGRSGRVGILFSHNR	60
	hap	1135	LDR+F + ++AVWT+ +D + Y S FRAY+Q+T+LRIG+QK L +GR+G +FSH+R	1194
60	orf1	61	TENTFDDGIGNSARLAHGAVFQGYGIDRFYXXXXXXXXXXXXXXXXXIGKXRRRVLHYG	120
	hap	1195	+NFTD+ + N A L + F QY K R+ + +G	1254
65	orf1	121	IQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPLGAENRYRAGIKADYSFKPA	180
	hap	1255	+A Y+ G GI+P+G RYF+++ +Y+ E V + TP LAENRY AGI+ DY+F P	1314
70	orf1	181	QHSITPPYLSLSYTDASGKVRTRVNTVLAQDFGKTRSAEAGVNAEIKGFTLSLHAAA	240
	hap	1315	+IS+ PY ++Y D ++ V+T VN VL Q FG+ E G+ AEI F+S + +	1374
	orf1	241	KGPOLEAQSAGIKLGYRW 259	
	hap	1375	+G QL Q +G+KLGYRW	1393

Homology with a predicted ORF from *N.gonorrhoeae*

The blocks of ORF1 show 83.5%, 88.3%, and 97.7% identities in 467, 298, and 259 aa overlap, respectively with a predicted ORF (ORF1ng) from *N.gonorrhoeae*:

5	orf1.pep	MKTTDKRTTETHRKAPKTGRIRFXAAYLAICLSGILPQAWAGHTYFGINYQYRDFAEN	60
	orf1ng	MKTTDKRTTETHRKAPKTGRIRFSAYLAICLSGILPQARAGHTYFGINYQYRDFAEN	60
10	orf1.pep	KGKFAVGAKDIEVYNNKKGLVKGSMTKAPMIDFSVSRNGVAALVGQYIVSVAHNGGYN	120
	orf1ng	KGKFAVGAKDIEVYNNKKGLVKGSMTKAPMIDFSVSRNGVAALAGDQYIVSVAHNGGYN	120
15	orf1.pep	NVDFGAEGNIXDQXRTYKIVKRNNYKAGTKGHPYGGDYHMPRLKHXVTDAPVPMTSY	180
	orf1ng	NVDFGAEGSN-PDQHRFSYQIVKRNNYKAGTNGHPYGGDYHMPRLKHVTDAPVPMTSY	179
20	orf1.pep	MDGRKYIDQNNYPDRVRIGAGRQYWRSDDEPNNRESSYHIAS-----	223
	orf1ng	MDGWKYADLNKYPDRVRIGAGRQYWRSDDEPNNRESSYHIASAYSWLVGGNTFAQNGSG	239
25	orf1.pep	-----GSPMFIYDAQKQKWLINGVLGTGPNFYLGKSGNG	255
	orf1ng	GGTVNLGSEKIKHSPYGFLLPTGSGFSGSPMFIYDAQKQKWLINGVLGTGPNFYLGKSGNG	289
30	orf1.pep	FQLVRKDWFYDEIFAGDTHSVFYEPFRNGKYSFNDDNNGTGKINAKHEHNSLPNLRKTRT	315
	orf1ng	FQLVRKDWFYDEIFAGDTHSVFYEPFRNGKYSFNDDNNGAGKIDAKHSHVSLPYLRKTRT	359
35	orf1.pep	VQLFNVLSETAREPVYHAAGGVNSYRPLNNGENISFIDECKGELILTSNINOAGAGLY	375
	orf1ng	VQLFNVLSETAREPVYHAAGGVNSYRPLNNGENISFIDCKGELILTSNINOAGAGLY	375
40	orf1.pep	FQGDFTVSPENNETWQAGVHISDSTVTWKVNGVANDRLSKIGKGT	422
	orf1ng	FEGNFTVSPKNETWQAGVHISDSTVTWKVNGVANDRLSKIGKGTLLVQAKENGQSGV	479
45	orf1.pep	DKVTASLTKTDISGNVDLADHAHLNLTGLA	744
	orf1ng	FGVAPHQSHICTRS DWTGLTSCTEKTITDDKVIASLKT DVRGNVSLADHAHLNLTGLA	774
50	orf1.pep	TLNGNLSANGDTR-YTVSHNATQNGXSLVXNAQATFNQATLNGNTSASGNASFNLSDBA	803
	orf1ng	TFNGNL-VQAEFTIRILRANATQNGNLSLVGNAQATFNQATLNGNTSASGNASFNLSDBA	833
55	orf1.pep	VQNGSLTSGNAKANVSHSALNGNVSLADKAVHFESRFTGQISGGDKTALHKLKDEWT	863
	orf1ng	VQNGSLTSDNAKANVSHSALNGNVSLADKAVHFENSRFTGKISGGDKTALHKLKDEWT	893
60	orf1.pep	LPSGTELGNLNDNATITLNSAYRHDAAAGTGSATDAPRRRSRRSRLXVTPPTSV	923
	orf1ng	LPSGTELGNLNDNATITLNSAYRHDAAAGTGSAAADAPRRRSRRS---LLSVTPPTSAE	950
65	orf1.pep	SRFNTLTVNGKLNQGTFRFMSELFGYRSGLKLAESSEGTYYLVANNNTGNEPVSLEQLT	983
	orf1ng	SRFNTLTVNGKLNQGTFRFMSELFGYRSGLKLAESSEGTYYLVANNNTGNEPVSLEQLT	1010
70	orf1.pep	VVEGDKNKLPSLENINFTLQNEHVDAGAW	1011
	orf1ng	VVEGDKNTPLSENINFTLQNEHVDAGAWRYQLIRKDGFRHLNPNVKEQELS DKLKAGET	1070
75	orf1.pep	LDRVFAEDRRNAVVTSGIRDTKHYRSQDFR	1211
	orf1ng	PQRDLISRYANSGLSEFSATLNSVFAVQDELDRVFAEDRRNAVVTSGIRDTKHYRSQDFR	1239
80	orf1.pep	AYRQQTDLRQIGMOKNLGSGRGVILFSHNRNTENTFDDGIGNSARLAHGAVFQGYIGDRFY	1271
	orf1ng	AYRQQTDLRQIGMOKNLGSGRGVILFSHNRNTENTFDDGIGNSARLAHGAVFQGYIGDRFY	1299

-372-

		10	20	30	40	50	60
	orf1-1.pep	MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGINYYQYRDFAE					
5	orf1ng-1	MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAAGHTYFGINYYQYRDFAE					
		10	20	30	40	50	60
	orf1-1.pep	70	80	90	100	110	120
10	orf1ng-1	KGKFAVGAKDIEVYNKKGELVGKSMTKA MIDFVSVRNGVAALVGDQYIVSVAHNGGYN					
		70	80	90	100	110	120
	orf1-1.pep	130	140	150	160	170	180
15	orf1ng-1	NVDFGAEGRNPDQHRFTYKIVKRNNYKAGTKGHFPGGDYHMPRLHKFVTDAEPVEMTSYM					
		130	140	150	160	170	180
	orf1-1.pep	190	200	210	220	230	240
20	orf1ng-1	DGRKYIDQNNYPDRVRIGAGRQYWRSD EDEPNNRESSYHLASAYSWLVGGTFAQNGSGG					
		190	200	210	220	230	240
	orf1-1.pep	250	260	270	280	290	300
25	orf1ng-1	GTVNLGSEKIKHSPYGF PTGGSGDGGSPMFIYDAQKQKWLINGVLGTGNFYIGKNSGF					
		250	260	270	280	290	300
	orf1-1.pep	310	320	330	340	350	360
30	orf1ng-1	QLVRKDWFYDEIFAGDTHSVFYEPQNGKYSFND DNGTGKINAKHEHNSL PNRLKTRTV					
		310	320	330	340	350	360
	orf1-1.pep	370	380	390	400	410	420
35	orf1ng-1	QLFNVSLSETAREPVYHAAGGVNSYR PRLNNGENISFIDEGKGELL L SNINQAGAGGLYF					
		370	380	390	400	410	420
	orf1-1.pep	430	440	450	460	470	480
40	orf1ng-1	QGDFTVSPENN ETWQAGVHISEDSTV TWKVG VANDRLSKIGKGL LHVQAKENQGSIS					
		430	440	450	460	470	480
	orf1-1.pep	490	500	510	520	530	540
45	orf1ng-1	EGNFTVSPK NETWQAGVHISEDSTV TWKVG VANDRLSKIGKGL LHVQAKENQGSIS					
		490	500	510	520	530	540
	orf1-1.pep	550	560	570	580	590	600
50	orf1ng-1	VGDGTVILDQ ADDKGGK AFSEIGLVSGRGTVQLNADNQFN PKLYYGF FRGGRLD NGH					
		550	560	570	580	590	600
	orf1-1.pep	610	620	630	640	650	660
55	orf1ng-1	SLSFHRIQNTDEGAMIVNHNQDK ESTV ITGNKDIAT TGN NNSLDSKKEIAYNGWFG EKD					
		610	620	630	640	650	660
	orf1-1.pep	670	680	690	700	710	720
60	orf1ng-1	MEGIPQGEI VWDNDWIDRT FKAENFHIQGGQAVVSRNVAKVGDWHL SNHAQAVGV VAPH					
		670	680	690	700	710	720
	orf1-1.pep	670	680	690	700	710	720
65	orf1ng-1	MEGIPQGEI VWDNDWIDRT FKAENFHIQGGQAVVSRNVAKVGDWHL SNHAQAVGV VAPH					
		670	680	690	700	710	720

730 740 750 760 770 780
orfl-1.pep QSHTICTRSDWTGLTNCVEKTIITDDKVIASLTKTDISGMVDLADHAHLNLTGLATLNGNL
 ||||| ||||| ||||| |||||
orflng-1 QSHTICTRSDWTGLTSCTEKTIITDDKVITASLKTDIRGNSVLADHAHLNLTGLATLNGNL
 ||||| ||||| ||||| |||||

10

[illegible]

15

850 860 870 880 890 900
 orf1-1.pep LSGNAKANVNSHSAINGVNSLADKAVFHFESSRFTGQISGGKDTALHLKNSWTLPSGSTEL
 |||||
 |||||
 orfing-1 LSGNAKANVNSHSAINGVNSLADKAVFHFESSRFTGQISGGKDTALHLKNSWTLPSGSTEL
 |||||
 |||||

20

```

          910      920      930      940      950      960
orf1-1.pep  GNLNLDNATITLNSAYRHDAAGAGTGSATDAPRRSRRRSRRLSVTPPTSVESRENTIT
             |||||
orfing-1    GNLNLDNATITLNSAYRHDAAGAGTGSADAPRRSR--RSLSVTPTTSAESRENTIT
             |||||
          910      920      930      940      950

```

25

```

          970      980      990      1000      1010      1020
orf1-1.pep  VNGKLGNGQSTFRFMSLLFGYRSDKLKLAESEGGTYTLVANNNGNEPASLEQLTVVEGKDN
          960      970      980      990      1000      1010
orf1ng-1    VNGKLGNGQSTFRFMSLLFGYRSGLKLAESEGGTYTLVANNNGNEPVSLEQLTVVEGKDN

```

30

```

              1030      1040      1050      1060      1070
orf1-1.pep    KPLSENINFTLQNEHVDAGAWRYQLIRKDGEFRLHNPVKEQELSKLGKA-----
              |||||
orf1ng-1      TPLSENINFTLQNEHVDAGAWRYQLIRKDGEFRLHNPVKEQELSKLGKAGETEAAALTA

```

35

1020	1030	1040	1050	1060	1070
------	------	------	------	------	------

1080 1090 1100 1110 1120
 orf1-1.pep ----EAKKQAEKDNQSLDALIAAGRDAVEKTESVAEPARQAGGENVGIMQAESEKKRVQ
 orf1ng-1 QAQLAAKQAEKDNQSLDALIAAGRNATEKAEVAEPARQAGGENAGIMQAESEKKRVQ
 1080 1090 1100 1110 1120 1130

40

```

1130      1140      1150      1160      1170      1180
orf1-1.pep ADKDTALAKQREARETPATTAFFPRARRRRDLPLQPQPQPQPQDRLISRYANGSLSEFS
orf1ng-1    ADKDTALAKQREARETPATTAFFPRARRRRDLPLQPQPQPQPQDRLISRYANGSLSEFS
1140      1150      1160      1170      1180      1190

```

15

	1190	1200	1210	1220	1230	1240
orf1-1.pep	ATLNSVFAVQDELDRVFAEDRRNAVVTSGIRDTHKHYRSQDFRAYRQOTDLRIGMQKNLIG					
orfing-1	ATLNSVFAVQDELDRVFAEDRRNAVVTSGIRDTHKHYRSQDFRAYRQOTDLRIGMQKNLIG					
	1200	1210	1220	1230	1240	1250

55

```

1250      1260      1270      1280      1290      1300
orf1-1.pep  SGRVGILFSHNRTNTFFDDGIGNSARLAHGAVFQYQIGRIDRYIIGISAGAGFSSGSLSDGI
|||||
orf1ng-1    SGRVGILFSHNRTGNTFFDDGIGNSARLAHGAVFQYQIGRIDRYIIGISAGAGFSSGSLSDGI
1260      1270      1280      1290      1300      1310

```

60

1310 1320 1330 1340 1350 1360
orf1-1.pep GGKIRRRVLHYGIQARYRAGPGFGIEPHIGATRYFVQKADYRYENVNVIATPGLAFNRYR
|||
orf1ng-1 RGKIRRRVLHYGIQARYRAGPGFGIEPHIGATRYFVQKADYRYENVNVIATPGLAFNRYR
1320 1330 1340 1350 1360 1370

22

```

1370      1380      1390      1400      1410      1420
orf1-1.pep AGIKADYSFKPAQHISITPYLSLSYTDAAAGKQVTRVNTAVLAQDFGKTRSAEWGVNAEI
|||||
orf1ng-1    AGIKADYSFKPAQHISITPYLSLSYTDAAAGKQVTRVNTAVLAQDFGKTRSAEWGVNAEI
1380      1390      1400      1410      1420      1430

```

5

10

15

20

25

30

35

40

45

50

55

60

65

70

		600	610	620	630	640	650
	orfing-1.pep	EKDATKTNGRINLNYQPEEADRTLLSSGNTLGNITQTNGLFFSGPTTHAYNHLGSGG					
5	p45387	ETDKNKHNGRLNLIYKPTTDRLLSSGNTLKGDIOTKGLFFSGPTTHAYNHLNKR					
		570	580	590	600	610	620
	orfing-1.pep	660	670	680	690	700	710
10	p45387	WSKMEGI PQGEIVWDNDWIDRTFKAENFHIQGGQAVVSRNVAKVEGDWHLNHAQVFGV					
		630	640	650	660	670	680
	orfing-1.pep	720	730	740	750	760	770
15	p45387	APHQSHSTICTRSNDWTGLTSCTEKITDDKVIASLKTDIRGNVSLADHAHLNLTGLATLN					
		690	700	710	720	730	740
	orfing-1.pep	780	790	800	810	820	830
20	p45387	GNLSAGSGDTHYTVRNATQNGNLSLVGNAQATFNQATLNGNTSASDNASFNLNNAVQNG					
		750				760	770
	orfing-1.pep	840	850	860	870	880	890
25	p45387	SLTSLDNKANKVSHSALNGNVSLADKAVHFHENSRTFGKISGGKDTALHLKDSWTLPSG					
		780	790	800	810	820	830
30	orfing-1.pep	900	910	920	930	940	950
35	p45387	TELGNLNLNDNATITINSAYRHDAAQAQTGSAADAPRRSRRLSLVTPPTSAESRFNLT					
		840	850	860	870		
	orfing-1.pep	960	970	980	990	1000	1010
40	p45387	VNGKLNQGTFRFMSELFGYRSGKLKLAESSEGTYTAVNNTGNEPVSLEQLTVVEGKDN					
		880	890	900	910	920	930
	orfing-1.pep	1020	1030	1040	1050	1060	1070
45	p45387	TPLSENLFNTLQNEHVDAGAWRYQLIRKDGFEFLHNPVKEQLSDKLKGAKGETAALTAK					
		940	950	960	970	980	990
	orfing-1.pep	1080	1090	1100	1110	1120	1130
50	p45387	QALAAKQAEKDNAQSLDALIAAGRNAT-EKAESVAEPARQAGGENAGINQAEAEKKRV					
		1000	1010	1020	1030	1040	1050
	orfing-1.pep	1140	1150	1160	1170	1180	1190
55	p45387	QADK---DTALAKQREATRATTAPFRRARRRD-LPQPOQPQPQQRDLISRYANGS					
		1060	1070	1080	1090	1100	1110
	orfing-1.pep	1200	1210	1220	1230	1240	1250
60	p45387	LSEFSAITNSVFAVQDELDRVFADERRNNAVWTSIGRTDKHYRSQDFRAYRQQTDLRQIG					
		1120	1130	1140	1150	1160	1170
	orfing-1.pep	1260	1270	1280	1290	1300	1310
65	p45387	MQKNLGSRGVILFSHNRNTGNTFDDGIGNSARLAHGAVFQYCGIGRFDIGISAGAGFSSG					
		1180	1190	1200	1210	1220	1230
70	p45387	VQKALANGRIGAVFSHRSRSDNTFDEQVKNHATLTMMSGFAQYQWGLDQFVNVGTGISAS					

-376-

		1320	1330	1340	1350	1360	1370
orfing-1.pep		SLSDGIRGKIRRRVLYHQIARYRAGGGFIEPHIGATRYFVQKADRYENVNIAIPGL					
5	p45387	1240	1250	1260	1270	1280	1290
		1380	1390	1400	1410	1420	1430
10	orfing-1.pep	AFNRYRAGIKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVLAQDFCKTRSAEW					
	p45387	1300	1310	1320	1330	1340	1350
		1440	1450	1460	1469		
15	orfing-1.pep	GVNAEIKGFTLSLHAAAAGPQLEAQHSAGIKGLGRWX					
	p45387	1360	1370	1380	1390		

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

20 Example 78

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 655>:

	1	..AAGGTGTGGC	AATTTGTCGA	AGA.CCGCTG	CGTGCCGTCG	TGCCTGCCGA
	51	CAGTTTGTAA	CGACCGCGCG	AAAAATTGAA	CCTGTTTAAG	GCGGGTGC
25	101	CAACCACTTT	GTTTTATGAA	GATCAAAATG	TCGTCAAAAG	TTTGCAGGAG
	151	CAGTTCCTCG	CTTATGCCCG	TAACTTCCCG	GTTCGGCGCG	ATCAGGCCAA
	201	CGCGATGGTG	CAGTATGCCG	TTTGGACGAC	ACITTCGCGC	GTCGGCGTAG
	251	GTGCAAACTC	GCAACATTAC	AATCCCTTGC	CCGATGCGCG	GATTGCCCAA
	301	GCGTGGGAATA	TCCCGGAAAA	CTGGTTGTGG	CGCGCACAAA	TGTTATCCGG
30	351	CGGTATTGAA	GGGGCGCGAC	GTGAAAAGAC	CTTTGAACCC	GTTCGAGAAC
	401	GTTCGAAAGT	GTTCGGCGCA	TAA		

This corresponds to the amino acid sequence <SEQ ID 656; ORF6>:

	1	..KVNQFVEXPL	RAVVPADSE	PTAOKLNLFK	AGAATILFEY	DQNVVKGLEQ
	51	QFPAYAAANP	VNADQANAMV	QYAVWTLAA	VGVGANLQHY	NPLPDAIAK
	101	AWNIPENWLL	RAQNVIGGIE	GAAGEKTFEP	VAERLKVFGA	*

35 Further sequence analysis revealed a further partial DNA sequence <SEQ ID 657>:

	1	..CTGCGTGCCG	TCGTGCCTGC	CGACAGTTTT	GAACCGACCG	CGCAAAAATT
	51	GAACCTGTTT	AAGGCGGGTG	CGGCAACCAT	TTGTTTAT	GAAGATCAAA
	101	ATGTCGTCAA	AGGTTTGCAG	GAGCAGTTCC	CTGCTTATGC	CGCTAACTTC
40	151	CCCGTTTGGG	CGGATCAGCG	AAACGCGATG	GTGCGATATG	CGGTTTGGAC
	201	GACACTTGCC	GCGGTGCGCG	TAGGTGCAAA	CCTGCAACAT	TACAATCCCT
	251	TGCCGAGATG	GCGGATTGCC	AAAGCGTGGA	ATATCCCCGA	AAACTGGTTG
	301	TTGCGCGCAC	AAATGGTTAT	CGGCGGTATT	GAAGGGGCGG	CAGGTGAAAA
	351	GACCTTTGAA	CCCGTTGCAG	AACGTTTGAA	AGTGTTCGCG	GCATAA

This corresponds to the amino acid sequence <SEQ ID 658; ORF6-1>:

	1	..LRVAVPADSF	EPTAQKLNLF	KAGAATILFY	EDQNVVKGLEQ	EQFPAYAAANP
45	51	PVWADQANAM	VQYAVWTLA	AVGVGANLQHY	NPLPDAIAK	KAWNIPENWLL
	101	LRAQNVIGGI	EGAAGEKTFE	PVAERLKVFG	A*	

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

50 ORF6 shows 98.6% identity over a 140aa overlap with an ORF (ORF6a) from strain A of *N.meningitidis*:

orf6.pep KVVQVFXEPLRAVVPADSFPEPTAQKLNLFK
 orf6a QIVEHAVLHTPSSFNSQSARVVVLFGEEHDKVWFQVEDALRAVVPADSFPEPTAQKLNLFK
 5 40 50 60 70 80 90
 orf6.pep AGAATILFYEDQNVKVGLQEQQFYAANFPFWADQANMVOYAVWTTTAAVGVGANLQHY
 orf6a AGAATILFYEDQNVKVGLQEQQFYAANFPFWADQANMVOYAVWTTTAAVGVGANLQHY
 10 100 110 120 130 140 150
 orf6.pep NPLPDAIAIAKAWNIPENWLLRQMVIGGIEGAAGKTFEPFAERLKVPGAX
 15 orf6a NPLPDAIAIAKAWNIPENWLLRQMVIGGIEGAAGKTFEPFAERLKVPGAX
 160 170 180 190 200

	1	ATGACGCTGCG	AATCTCTGCA	ACAGGTCGGCA	GAGAGCGCGC	GTTCGATTTA
20	51	TTGCTAAAT	AAAATCTGC	CCGCAATCGC	CGCTGAATCT	CCGCAATCTG
	101	TGCGACAGCG	CGTITGGAC	ACACCTTTCT	CGTTCAATTG	CCAACTTGCC
	151	CGTGTGTCG	TGCTGTTTG	GACGAGACAT	GAATAGGTGT	GCGCAATTGT
	201	CGAGACGCG	TCGCGTGGC	TCGTGCTCGT	CGACAGTTT	GAACGACGC
	251	CGCAAAATT	GAACTCTTT	AAGCGGGTG	CGCGAACTAT	TTGTGTTTAT
25	301	GAGATCAAA	ATGTGCTCA	AGSTTTCGAG	GAGCAATTC	CTGCTTATGC
	351	CGCACTTIT	CCGCTTTGG	GCGACCGAC	GACGCGATG	GTGCAGTATG
	401	CGGTTTGGAC	GACCTATGCG	CGGCTCGCG	TAGGTGCAAA	CTGCTCAACT
	451	TACATCCCT	TGCCGATGCT	GCGCAATCG	AAGGCTGAA	ATATCCGAA
	501	AACCTGGCG	CGCGGACG	AAATGGCTTT	CGGCTGCTG	CGGCTGCTG
	551	CAGTGAAAA	GACCTTTGAA	CGAATTGAG	AACGTTTGAA	AGTGTTCGCG
30		GATTA				

35

1	MTRQSLQQA	ESRRSIYSLN	KNLPVGKDEI	VQIVEHAVLH	TPSSFNSQSA
5	RVVVLFGEEH	DKVQFVEDA	LRAVPVADMF	EPTAQKNLNF	KAGAAATILFY
101	EDQNVVVGGLQ	EQFFVAAHFN	PVWADQADAM	VQYAVWTLTA	AVGGANGLQF
151	YNPLPDAATA	KAWNIPENWL	LRAQMVIIGI	EGAAGEKTFE	PVAERLKVFG
201	A*				

		50	60	70	80	90	100
40	orf6a.pep	TPSSFNQSARVVVLFGEEHDKVQFVEDLARAVVPDASEPTAQKLNLFKGAATILFY 					
	orf6-1	LRAVVPADSFEPTAQKLNLFKGAAITILFY 102030					
45		110	120	130	140	150	160
	orf6a.pep	EDQNNVKGLQEOPPAYAANFPWADQANAMVOYAVWTTLAAVGVGANLQHYNPLPDAATA 					
	orf6-1	EDQNNVKGLQEOPPAYAANFPWADQANAMVOYAVWTTLAAVGVGANLQHYNPLPDAATA 405060708090					
50		170	180	190	200		
	orf6a.pep	KAWNIPENWLLRAQMVIIGGEAGAGKETEEVAERLKVFAGX 					
	orf6-1	KAWNIPENWLLRAQMVIIGGEAGAGKETEEVAERLKVFAGX 100110120130					
55							

ORF6 shows 95.7% identity over a 140aa overlap with a predicted ORF (ORF6ng) from *N.gonorrhoeae*:

-378-

	orf6.pep	KVWQFVEKPLRAVVPADSFEP	TAQKLNLFK	30
	orf6.ng	SNVSLMSNPTVLRMGLFLYIASLRGAIYKVWQFVEDALRAVVPADSFEP	TAQKLELFK	64
5	orf6.pep	AGAATILFYEDQNVVKGLQEOPFAYAAANFPVWADQANAMVOYAVWTT	LAAVGVGANLQHY	90
	orf6.ng	AGAATILFYEDQNVVKGLQEOPFAYAAANFPVWADQANAMVOYAVWTT	LAAVGAGANLQHY	124
10	orf6.pep	NPLPDAAIAKAWNIPENWLLRAQMVGIGIEGAAGEKTFEP	VPAERLKVFGA	140
	orf6.ng	NPLPDVAIAKAWNIPENWLLRAQMVGIGIEGAAGEKVFEP	VPAERLKVFGA	174

The complete length ORF6ng nucleotide sequence <SEQ ID 661> was identified as:

15	1	ATGGCCGTTG	CGTCAAATGT	CAGCTTGGAT	ATGTCCAATC	CTACGGTGTT
	51	ACGCATGGGA	TTACCCCTAT	ATAITGCGTC	CCTAAGAAGG	GGCGCAATAT
	101	ATAAGGTGTG	GCAATTTGTC	GAAGACGGGC	TGCGTGGCGT	CGTGGCTGGC
	151	GACAGTTTGT	AACCGACCGC	GCAAAAATGT	AAGCTGTGTA	AGCGGGCGGC
	201	GGCAACCATT	TGTTTTATG	AGATCAAAA	TGTCGTCAAA	GGTTTCAGG
	251	ACGAGTTCCC	TGCTTATGCC	GCCACTTTC	CCGTTTGGGC	GGACCAGGGC
20	301	AACGCTATGG	TACAGTATGC	CGTCTGACG	ACACTTGGCG	CGGTCGGTGC
	351	AGGTGCAAA	CTGCAACATT	ACAACCCCTT	GCCCGATGTG	GCGATTGCTA
	401	AAGCGTGGAA	TATTCGCCAA	AACCTGGCTG	TGCGCGCGCA	AATGGTTATC
	451	GGTGGTATTG	AAGGGGcgcc	aggtgaaaaa	gtctttgaac	CCGTTGCgga
	501	acgtttgAAA	GTGTTGCGGC	CATAA		

25 This encodes a protein having amino acid sequence <SEQ ID 662>:

30	1	MAVASNSVLD	MSNPTVLRMG	LPLYIASLRR	GAIKYKVQFV	EDALRAVPA
	51	DSFEPTAQL	KLFKAGAATI	LFYEDQNVVK	GLOEOPFAYA	ANFPVWADQA
	101	NAMVQYAVMT	TLAAGVAGAN	LQHYNPLPDV	AIKAWNIP	NWLLRAQMVI
	151	GGIEGAAGEK	VFEPAERLK	VFGA*		

ORF6ng and ORF6-1 show 96.9% identity in 131 aa overlap:

	orf6-1.pep			10	20	30
				LR	AVVPADSFEP	TAQKLNLFKAGAA
35	orf6.ng	PTVLRMGLFLYIASLRGAIYKVWQFVEDALRAVVPADSFEP	TAQKLNLFKAGAA	10	20	30
		20	30	40	50	60
		40	50	60	70	80
40	orf6-1.pep	EDQNVVKGLQEOPFAYAAANFPVWADQANAMVOYAVWTT	LAAVGVGANLQHYNPLPDAAIA			
	orf6.ng	EDQNVVKGLQEOPFAYAAANFPVWADQANAMVOYAVWTT	LAAVGAGANLQHYNPLPDVAIA			
		80	90	100	110	120
		100	110	120	130	
45	orf6-1.pep	KAWNIPENWLLRAQMVGIGIEGAAGEKTFEP	VPAERLKVFGAX			
	orf6.ng	KAWNIPENWLLRAQMVGIGIEGAAGEKVFEP	VPAERLKVFGAX			
		140	150	160	170	

50 It is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 79

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 663>

55	1	..GGCTACAAC	ACCTGTTCCG	GCGCGGCAGC	CGCATCGCCA	ACTACCAAT
	51	CAACGGCAGC	CCCGTTGCCG	ACGCGCTGGC	CGATACGGGT	CAATGCCAAC
	101	ACCGCCGCTT	ATGAGCGCGT	AGAAGTCGTG	CGCGGCTGGG	CGGGCTGCTG
	151	GGACGGCAGC	GCGCAGCCTT	CCGCCACCGT	CAATCTGGTG	CGCAAAACGC
	201	TGACCCGCAA	GCCATTGTTT	GAACTCCGCG	CCGAGCGGGG	CAACCGGAAA

251 CATTTTCGGGC TGGACGCGGA CGTATCGGGC AGCCTGAACA CCGAAG. crC
 301 rCTGCGCGCG CGCTGGTTT CCACTTCGGC ACGCGCGGAC TCGTGGCGGC
 351 GGCGCGAAGC CAGCCGSKAT GCGGAACCTT ACGGCATTTT GGAATACGAC
 401 ATCGCACCGC AAACCCGCGT CACGCGAAGC ATGGACATC AGCAGCGGAA
 451 AGAARACCGC GACGCGCGCG TCAGCTACGC CCGTGTACGAC AGCCAGGTT
 501 ATGCCACCGC CTTCCGCGCG AAAGACAACG CGCCACACAA TTGGGCGGAT
 551 AGCCACACCC GTGCGCTCAA CCGTGTTCGC GGCATCGAAC ACCGCTTCAA
 601 CCAAGACTGG AAACCTCAAG CGAATACGA CTAC..

This corresponds to the amino acid sequence <SEQ ID 664; ORF23>:

10 1 ..GYNLFFARGS RIANYQINGI PVADALADTG NANTAAYERV EVVRGVAGLL
 51 DGTGEP SATV NLVRKRLTRK PLFEVRAEAG NRKHFGLDAD VSGSLNTEXX
 101 LRGRLVSTFG RGDWRRRER SRXAEYLGIL EYDIAPQTRV HAXMDYQOAK
 151 ETADAPLSYA VYDSQGYATA FGPKDNPATN WANSNHRALN LFAGIEHFRFN
 201 QDWKLEAYD Y..

15 Further work revealed the complete nucleotide sequence <SEQ ID 665>:

1 ATGACACGCT TCAAAATATC CCGCTGCTTT GCGCGCGCTT TGCCCGTGTA
 51 CGCGACGGCG GATGTTTTCTG TTTCAAGAGA CCCCACACCG CAGGAAGAGC
 101 CTGAATTTGCC GACCATCACC GTTACCGCGC ACCGACACCG GAGTTCCAAC
 151 GACGGCTACA CTGTTTTCCG CACGCAACCC CGCGTCGGCG TGCCCATGAC
 201 CCGCGCGGAA ATCCCGCAGA GCGTCAGCGT CATCACATCG CAACAATGCT
 251 GCGACCAAAA CATCAAAACG CTCGACCGCG CCGTGTGTGA GCGGACCGGC
 301 ACCAGCGCGC AGATTTACGG CTCGACCGCG GCGGGCTACA ACTACTGTTT
 351 CGCGCGCGCG AGCGCATCG CCAACTACCA AATCAACGCG ATCCCGGTTG
 401 CCGACGCGCT GCGCGATACG GGCATGCGCA ACACGCGCG CTATGAGCGC
 451 GTAGAAGTCG TGCGCGCGCT GCGGGGGCTG CTGGACGCGCA CGGCGAGCC
 501 TTCCGCGCAC GTCAATCTGG TGCGCAAAAG CCGTACCGCG AAGCCATTGT
 551 TTGAAGTCG CGCGGAGCG GGCACCGCGA AACATTTCGG GCTGGACGCG
 601 GACGTATCGG GCAGCCTGAA CACCGAAGCG ACGCTGCGCG GCGCGCTGGT
 651 TTCCACCTTC GGAAGCGCGG ACTCGTGGCG GCGCGCGGAA CGCAGCGCGG
 701 ATGCGCAACT CTACGCGATT TTGAATACG ACATGCGACC GCAAAACCGC
 751 GTCCAGCGCG CAGTCGATGCA CACGACGCGC AAGAAGACCG CCGACGCGCC
 801 GCTCAGCTAC GCGTGTACG ACACGCGAGG TTATGCGCAC GCGTTCGCGC
 851 CGAAGACTCA CCGCGCGACA AATTGCGGCA ACACGCGCGA CCGTGCCTCG
 901 AACCTGTTCG CCGCGATCGA ACACGCGTTC AACCAAGACT GGAACACTCA
 951 AGCCGAATAC GACTACACCC GCGCGCGCTT CCGCGACGCG TACGCGGTAG
 1001 CAGCGGTGCT TTCCATCGAC CACACACCG CGCGCACCGA CCGTATTCCG
 1051 GGTATTGGC ACGCGGACCC GCGCACCCAC AGCGCCAGCG GTGCTATTAG
 1101 CGGCAAAATAC CGCTGTTCG GCGCGGAACA CGATTTAATC CGGGGTATCA
 1151 ACGGTTACRA ATACGCGCAG AACAAATACG GCGACGCGAG CATCATCCCC
 1201 AAGCGCAATC CCAACGCGTA CGAATTTTCC CGCAGGGGTG CCTACCGCGA
 1251 GCGTGCATCG TTTGCCCAAA CCATCCGCGA ATACGCGACC AGGCGGCAAA
 1301 TCGCGCGCTA TCTCGCCACC CGTTCGCG CGCGCGACAA CCGTTCGCTG
 1351 ATTTTGGGCG GACGATACAC CCGTTACCGC ACGCGAGCT ACACGAGCGG
 1401 CACACAAGGC ATGACCTATG TGTCGCGCAA CGGTTTACCC CCTACACAGC
 1451 GCATCGTGTG CGACCTGACC GGCACCTGT CTCCTTACGG CTCGACAGCG
 1501 AGCCTGTTCG TCCCGCAATC GCACAAAGAC GAACACGCGA GCTACTGTAA
 1551 ACCCGTAACC GGCAACAAATC TGAAGCGCG CATCAAGGCG GAATGGCTTG
 1601 AAGCGCGCTC GAACGATACC GCGCGCGTGT ACGCGCGCG TAAACACAC
 1651 CTGCGCACCG CAGCAGACCG GACCGCGAGC GCGACACCT ACTACCGCGC
 1701 CGCCACCAA GCCAAACACC ACGCGTGGGA AATCGAAGCT GCGCGCGCGA
 1751 TCACGCGCGA ATGCGAGATA CAGGCGAGTT ACAGCGCAAG CAAACCCCGC
 1801 GACCAAGACG CGACGCGCTT GRACCGGAC AGGCTACCGC AACGCGCTT
 1851 CAACCTCTTC ACTGCTTACC ACTTTGCCCG CGAAGCCCCC ACGCGCTGGA
 1901 CCATCGGCGC AGCGTGTGCG TGCGAGAGCG AAACCCACAC CGACCTGCGC
 1951 ACGCTTCGCA TCCCAACACC CGCGCGCAAA GCGCGCGCGC CGACACAGC
 2001 CGCGCAAAAA GCCTACGCGC TCGCGACAT CATGGCGGT TACCGCTTCA
 2051 ATCCGCGCGC CGAAGCTGCG CTGAACGTGG ACAATCTGT CACAAACAC
 2101 TACCGCACCC AGCCGACCGC CCACAGCTAC GCGCACTGCG GGACAGTGAA
 2151 CGCGCGGTTT ACCTATCGGT TTAAATA

60 This corresponds to the amino acid sequence <SEQ ID 666; ORF23-1>:

1 MTRFKYSLLF AALLPVYAQA DVSVDPKPK QESTLEPTIT VTADRIASSN
 51 DGYTVSGTHT PIGLPMTLRE IPQSVSVITS QMRDQNKIT LDRALLQATG
 101 TSQRTYGS DR AGYNLFARG SRIANYQING IPVADALADT GNANTAYER
 151 EVVRGVAGL LDGTGEP SAT VNLVRKRLTR KPLFEVRAE NRKHFGLDA
 201 DVSGSLNTEG TLRGRLVSTF GRGDSWRKRE RSRDAELYGI LEYDIAPQTR

-380-

251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNFAT NWANSRHRAL
 301 NLFAGIEHRF NODWKLKAEY DYTRSRFRQP YGVAGVLSID HNTAATDLIP
 351 GYWHADPRTH SASVSLIGKY RLFGREHDLI AGINGYKYAS NKYGERSIIP
 401 NAI PNAYEFS RTGAYPQPAS FAQTIPQVGT RRQIGGYLAT RFRAADNLSL
 451 ILGGRYTRYR TGSYDSRTQG MTYVSANRFT PYTGIVFDLT GNLSLYGSYS
 501 SLFPVQSQKD EHGSYLKPVT GNNLEAGIKG EWLEGRINAS AAVYRARKNN
 551 LATAAGRDPS GNTYRAAANQ AKTHGEWIEV GGRITPEWQI QAGYSQSKTR
 601 DQDGSRLNPD SVPEPSFKLF TAYHFAPEAP SGWTIGAGVR WQSETHDTPA
 651 TLRIPNPAK ARAADNSRQK AYAVADIMAR YRFNPRAEALS LNVDNLFNKH
 701 YRTQPDHRSY GALRTVNAAF TYRFK*

Computer analysis of this amino acid sequence gave the following results:

Homology with the ferric-pseudobactin receptor PupB of *Pseudomonas putida* (accession number P38047)

ORF23 and PupB protein show 32% aa identity in 205aa overlap:

Orf23 6 FARGSR IANYQINGIPVADALADTGNANTAA YERVEVVRGVAGLLDGTGEPSATVNLVRK 65
 ++RG I NY++G+P + L D + + A ++RVE+VRG GL+ G G PSAT+NL+RK
 PupB 215 WSRGFAIQNYEVDGVPTSTRL-DNYSQSNAMFDRVEIVRGATGLISGMGNFSAITINLIRK 273
 Orf23 66 RLTRKPLFEVRAEAGNRKHFGLDADVSGSLNTEXXLRGLRVSTFXXXXXXXXXXXXXXAE 125
 R T + + EAGN -G DVSG L +RGR V+ +
 PupB 274 RPTAEQAQASITGEAGNWDRYGTGFVSGFLTETGNIRGRFVADYKTEKAWIDRYNQOSQL 333
 Orf23 126 LYGILEYDIAPQTRVHAXMDYQQA KETADAPLSYAVYD--SQGYATAFGPKDNPATNWAN 183
 +YGI E+D++ T + Y + D+EL + S G T + N A +R+
 PupB 334 MYGITFEDLSEDTLLTVGFSY--LRSDISFLRSGLPTRFSTGERTNLKRSINAAFDWSY 391
 Orf23 184 SHHRAINLFAGIEHRFNQDWKLKAE 208
 + H + F IE + W K E
 PupB 392 NDHEQTSFFTSIEQQLGNGWSGKIE 416

30 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF23 shows 95.7% identity over a 211aa overlap with an ORF (ORF23a) from strain A of *N.*

meningitidis:

orf23.pep 10 20 30
 GYNLFAFGSR IANYQINGIPVADALADTG
 35 orf23a QMRDQNIKALDRALLQATGTSRQIYGSDRAGYNLFAFGSR IANYQINGIPVADALADTG
 90 100 110 120 130 140
 orf23.pep 40 50 60 70 80 90
 NANTAA YERVEVVRGVAGLLDGTGEPSATVNLVRKRLKPLFEVRAEAGNRKHFGLDAD
 orf23a NANTAA YERVEVVRGVAGLLDGTGEPSATVNLVRKRLKPLFEVRAEAGNRKHFGLDAD
 150 160 170 180 190 200
 orf23.pep 100 110 120 130 140 150
 VSGSLNTEXXLRGLRVSTFGRGDSWRRRERSRKAELYGILEYDIAPQTRVHAXMDYQQA
 orf23a VSGSLNTEXXLRGLRVSTFGRGDSWRRRERSRKAELYGILEYDIAPQTRVHAXMDYQQA
 210 220 230 240 250 260
 orf23.pep 160 170 180 190 200 210
 ETADAPLSYAVYDSQGYATAFGPKDNPATNWANSRHRALNLFAGIEHRFNQDWKLKAEYD
 orf23a ETADAPLSYAVYDSQGYATAFGPKDNPATNWANSRHRALNLFAGIEHRFNQDWKLKAEYD
 270 280 290 300 310 320
 orf23.pep Y
 I
 60 orf23a YTRSRFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTHASVSLIGKYRLFGRHDLIA
 330 340 350 360 370 380

The complete length ORF23a nucleotide sequence <SEQ ID 667> is:

1 ATGACACGCT TCAAAATATTC CTTGCTGTTT GCCGCCCTGT TGCCCGTGTA
 51 CGCGCAGGCC GATGTTTCTG TTTACAGACA CCCAAACCG CAGGAAGACA
 101 CTGAATTGCC GACCATCACG GTTACGCGCG ACCGCACCGG GAGTTCACAC
 151 GAGCGGTACA CTGTTTCCGG CAGCACAACG CGGCTCGGGC TGCCCATGAC
 201 CCGTGGCGAA ATCCCGCAGA GCGTCAGCGT CATCATCATG CAACAATATG
 251 GCGACCAAAA CATCAAGAGG CTCGACGCGG CCTGTGTGCA GCGACGCGCG
 301 ACCAGCGCGC AGATTTCAGG CTCGACGCGG GGGGGCTACA ACTACTGTTT
 351 CGCGCGCGGC AGCGCGATCG CCAACTACCA AATCAACGGC ATCCCGGTGT
 401 CGACGCGCGT GCGCGATACG GGCATGCGCA ACACGCGCGC CTATGAGGCG
 451 GTAGAAGTCG TGCGCGGGGT GCGCGGGCTG CTGGACGGCA CGGCGAGGCC
 501 TTCCGCGCAC GTCAATCTGG TGCGCAAAAG CCGGACCGCG AAGCGATTGT
 551 TTGAAGTCGG CGCGCAAGGG GCGCAACGCA AHCATTTCGG GCTGGGGGCG
 601 AGCGTATCGG GCAGCGCTGAA TCCCGAAGCG ACAGTGGCGG CGGCGGTGTG
 651 TTCCACCTTC GAGCGCGGG ACTCGTGGCG GAGCGCGCAA CGCAGCGCGG
 701 ATCCGGAATC CTACGCGCAT TTGGAATACG ACATGCAAC GMAAACCGCG
 751 GTCCAGCGAG GCATGGGACT CACGACGGCG AAGAAGACCG CCGACGCGCG
 801 GCTCAGCTAC GCGGTGTACG ACAGCGAAGG TTATGCCAAC CGCTTCGGCC
 851 CGAAAGACAA CCGCGCCACA AATTGGGCGA ACAGCGCGCA CGGTGGCGCT
 901 AACCTGTTTG CGCGCATCGA ACACGCGCTT AACCAAGACT GGAATCTCAA
 951 AGCGCAATAC GACTACACCC GACGCGGCTT CCGCGACGCC TACGGCGTAG
 1001 CAGGCGTGCT TTCCATCGAG CACAACACCG CGCGCACCGA CTGTATTCCC
 1051 GGTATTGGCG ACGCGAGCCG GCGCACCCAC AGCGCGAGCG TGTCATTAAAT
 1101 CGGCAATATC CGCTGTTTGG CGCGGGAACA CGATTTAATC GCGGGTATCA
 1151 ACGTTACAAA ATACGCGCAGC AACAAATACG GCGAAGCGAG CATCATCCCC
 1201 AACGCGATTG CCAACGGCTA GGAATTTTCC CGACGCGGTG CTAACCGCGA
 1251 GCGTGCATCG TTTGCCGAAA CCATCCGCGA ATAAGGCAAC AGGCGGCAAA
 1301 TCGGCGGCTA TCTCGCCACC CGTTTCGCGG CCGCGACAAA CTTTTCGCTG
 1351 ATACTCGGGG GCAGATACAG CCGTTACGCG ACCGCGAGCT ACAGACGCGG
 1401 CACAGAAGCG ATGACCTATG TGTCCGCGAA CGCTTTCACC CCTACACAG
 1451 GCATCGTGTT CGACCTGACC GGCACCTGT CGCTTACCG CTGCTACAGC
 1501 AGCGCTGTGG TCCCGCAATC GCAAAAGACG GACACGGCGA GCTACTGAA
 1551 ACCCGTAACC GCGCAACATC TGAAGCGCGG CATCAAGGCG GAATGGCTTG
 1601 AAGCGCGTCT GRACGCAATC GCGCGCGTGT ACCGCGCGCG TAAAAACAC
 1651 CTCGCCACCG CAGCAGGACG GSACCGCGAGC GGCACACCT ACTACGCGCG
 1701 CGCCAAACCA GCCAAAACCC ACGCGTGGGA AATCGAAGTC GCGCGCGCGA
 1751 TCAAGCCCGA ATGGCAGATA CAGGACAGTT ACAGGCCAAG CAAAACCGCG
 1801 GACCAAGAGC GCAGCGCGCT GAAACCGGAC AGCGTACCGG AACGCGAGTT
 1851 CAACTCTTTC ACTGCGCTACC ACTTTGCGCC CGAAGCGCGC AGCGGTGGA
 1901 CCATCGCGCG AGCGGTGGCG TGGCAGAGCG AAACCCACAC CGACCTGCC
 1951 ACGCTCCGCA TCCCAACCC CGCGCGCAA GCGCGCGCG CCGACAACAG
 2001 CCGCCAAAAA GCCTACGCGG TCGCGGACAT CATGGCGCGT TACCGCTTCA
 2051 ATCCGCGCGC CGAAGTGTGG CTGAAGCTGG ACAATCTGTT CAACAACAC
 2101 TACGCAACCC AGCCGACCG CCAAGCTATC GCGCGACTGC GGACAGTGAA
 2151 CGCGCGGTTT ACCTATCGGT TTAATAA

45 This encodes a protein having amino acid sequence <SEQ ID 668>:

1 MTRFKYSLLE AALLPVYAQA DVSVSDPKP QESTELPTIT VTADRTASSN
 51 DQTVSGTET PLGLEMLIRE IQQSVSVITS QQMRDNKIA LDRALLQATG
 101 TSROIYSDSR AGYNYLFARG SRIANYOING IPVADALADT GNANTAYER
 151 VEVVRGVAGL LDGTGEPSTAT VNLVRKRPRTR KPLFEVRAEA GNRKHFGLGA
 201 DVSGSINAEG TLRGRIVSTF GRGDSWRQRE RSRDAELYGI LEYDIAFOTR
 251 VHAGMDYQQA KETADAPLSY AVYDSOGVAT AFPGKCNFAT NWNASRHRAL
 301 NLFAGIEHRF NODWKLKAEY DYTRSRFRQP YGVGVLISID HNTAATDLIP
 351 GYWHADPRTH SASVSLIGKY RLFGRHDLII AGINGYKYAS NKYGERSIIP
 401 NAIPNAYEFS RTGAYPOPAS FAQTIPOYGT RRQIGGYLAT RFRAADNLISL
 451 ILGGRYSRYR TGSYDSRTQG MTYVSANRFT PFTGIVFDLT GNLSLYGSYS
 501 SLFVPQSQKD EHGYSLEKPV TGNLEAGIKG EWLEGRINAS AAVYRARKNN
 551 LATAGRDRPS GNTYRYAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKTR
 601 DQDGSRLNPD SVPSERFKLF TAYHFAPEAP SGWITIGAGVR WQSEHTDTPA
 651 TLRIPNPAK ARAADNRQK AYAVADIMAR YRPNPRAELS LNVNLEFNKH
 701 YRTQDRHSY GALRTVNAAF TYRFK*

ORF23a and ORF23-1 show 99.2% identity in 725 aa overlap:

		10	20	30	40	50	60
orf23a.pep	MTRFKYSLLEAALLPVYAQA	DVSVSDPKPQESTELPTIT	VTADRTASSN	DGYTVSGTET	PLGLEMLIRE	IQQSVSVITS	QQMRDNKIALDRALLQATG
65 orf23-1	MTRFKYSLLEAALLPVYAQA	DVSVSDPKPQESTELPTIT	VTADRTASSN	DGYTVSGTET	PLGLEMLIRE	IQQSVSVITS	QQMRDNKIALDRALLQATG
	10	20	30	40	50	60	

		70	80	90	100	110	120
	orf23a.pep	PLGLPMTLREIPQSVSVITSQQMRDQNIKALDRALLQATGTSRQIVGSDRAGYNYLFARG					
5	orf23-1	PLGLPMTLREIPQSVSVITSQQMRDQNIKTLDRALLQATGTSRQIVGSDRAGYNYLFARG					
		70	80	90	100	110	120
	orf23a.pep	130	140	150	160	170	180
10	orf23-1	SRIANYQINGIPVADALADTGNANTAAERVEVVRGVAGLLDGTGEPASATVNLVRKKRLTR					
		130	140	150	160	170	180
	orf23a.pep	190	200	210	220	230	240
15	orf23-1	KPLFEVRAEAGNRKHFGLGADVSGSINAGTLLRGLRVSTFGRGDSWRQRERSRDAELYGI					
		190	200	210	220	230	240
	orf23a.pep	250	260	270	280	290	300
20	orf23-1	LEYDIAPIQTRVHAGMDYQAKETADAPLSYAVYDSQGYATAFGPKONPATNWNASRRHAL					
		250	260	270	280	290	300
	orf23a.pep	310	320	330	340	350	360
25	orf23-1	NLFAGIEHRFNQDWKLEAYDYTRSRFRQPYGVAGVLSIDHNNTAATDLIPGYWHADERTH					
		310	320	330	340	350	360
	orf23a.pep	370	380	390	400	410	420
30	orf23-1	SASVSLIGKYRLFGREHDLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYPQPAS					
		370	380	390	400	410	420
	orf23a.pep	430	440	450	460	470	480
35	orf23-1	FAQTIPQYGTTRQIGGYLATRFRAADNLSLILGGRYTRYRTGSYDSRTQGMTYVSANRFT					
		430	440	450	460	470	480
	orf23a.pep	490	500	510	520	530	540
40	orf23-1	PYTGIVFDLTGNLSLYGSYSSLFVPSQSKDEHGSYLPVTGNNEAGSKGEWLEGRINAS					
		490	500	510	520	530	540
	orf23a.pep	550	560	570	580	590	600
45	orf23-1	AAVYRARKNNLATAAGRDPSGNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSKTR					
		550	560	570	580	590	600
	orf23a.pep	610	620	630	640	650	660
50	orf23-1	DQDGSRLNPDSVPERSFKLFTAYHFAPEAPSGWTIGAGVWQSETHTPATLRIENPAK					
		610	620	630	640	650	660
	orf23a.pep	670	680	690	700	710	720
55	orf23-1	ARAADNSROKAYAVADIMARYRENPAELSLNVDNLFNKHYRTQPRHSYGALRTVNAAF					
		670	680	690	700	710	720
	orf23a.pep	TYRFXK					
60	orf23-1	TYRFXK					

Homology with a predicted ORF from *N.gonorrhoeae*

ORF23 shows 93.4% identity over a 211aa overlap with a predicted ORF (ORF23.ng) from *N.*

gonorrhoeae:

5	orf23.pep	GYNYLFARGSR IANYQINGIPVADALADTGNANTAAERVEVVRGAGLLD	51
	orf23.ng	SAVDACRIPGYNYLFARGSR IANYQINGIPVADALADTGNANTAAERVEVVRGAGLLD	60
	orf23.pep	GTGEP SATVNLVRKRLTRKPLFEVRAEAGNRKHFLDADVSGSLNTEXXLRGLVSTFGR	111
10	orf23.ng	GTGEP SATVNLVRKHPTRKPLFEVRAEAGNRKHFLGADVSGSLNAEGTLRGLVSTFGR	120
	orf23.pep	GDSWRRRERSRXAELYGILEYDIAPOTRVHAXMDYQOAKETADAPLSYAVYDSQGYATAF	171
	orf23.ng	GDSWRQLERSRDAELYGILEYDIAPOTRVHAGMDYQOAKETADAPLSYAVYDSQGYATAF	180
15	orf23.pep	GPKDNPATNWANSHHRLNLFAGIEHRFNQDWLKAEDY	211
	orf23.ng	GPKDNPATNWSNRNALNLFAGIEHRFNQDWLKAEDYDTSRSRFPQYGVAGVLSIDHS	240

The ORF23ng nucleotide sequence <SEQ ID 669> is predicted to encode a protein comprising

20 amino acid sequence <SEQ ID 670>:

1	SAVDACRIPG	YNYLFARGSR	IANYQINGIP	VADALADTGN	ANTAAYERVE
51	VVRGAGVLPD	GTGEP SATVN	LVRKHPTKRF	LFEVRAEAGN	RKHFLGADV
101	SGSLNAEGTL	RGRVSTFGR	GDSWRQLERS	RDAELYGILE	YDIAPOTRVH
151	AGMDYQOAKE	TADAPLSYAV	YDSQGYATAF	GPKDNPATNW	SNRSNRNALN
201	FAGIEHRRFNQ	DWLKAEDYD	TSRSRFPQY	VAGVLSIDHS	TAATDLIPGY
251	WHADPRTHSA	SHSLTKGYRL	FGREHDLIAG	INGYKYASNK	YGERSIIPNA
301	IPNAYEFSRT	GAYEPQSSFA	QTIPQYDTRR	QIGGYLATRF	RAADNLSLIL
351	CGSRSPQMG	TVSANRTPFY	TSVISLTLGN	LSLYGKQVGE	YVRAKKNWLA
401	FVPLQKDEH	GSYLKIPVTGN	MLEADIKGEW	LEGRINASAA	VYRANKKNWLA
451	TAAGRDQSGN	TYVRAANOAK	THGWIEIVCG	RITPEWIOQA	QYSQSKPRDQ
501	DGSRLPNDPS	PERSFKLFTA	YHLAPEAPSG	RTIGAGVVRQ	GETHTDPAAL
551	RIPNPAAKAR	AVANSRQKAY	AVADIMARYR	FNPRTELSLN	VDNLNKHRYR
601	TQDPRHSYGA	LRTVNAAFY	RFK*		

Further work revealed the complete nucleotide sequence <SEQ ID 671>:

35	1	ATGACACGCT	TCAAATACTC	CCTGCTTTT	GCGCGCTGC	TACCGGTGA
	51	CGGCGAGGCC	GATGTTTCTG	TTTCAGACGA	CCCCAAACCG	CAGGAAAGCA
	101	CGGAATTGCC	GACCATCAC	GTTACCGCG	ACCGCACCGC	GAGTTCACAC
	151	GACGGCTACA	CCGTTTCGGG	CACGCACACC	CCGTTCCGGG	TGCCCATGAC
	201	CCTGCGCGAA	ATCCGCGAGA	GGCTCAGCT	CATCACATCG	CAACAATGCG
40	251	GCGACCAAAA	CATCAAAACG	CTCGACCGCG	CCCTGTGGCA	GCGACCGCGC
	301	ACCGACCGCG	AGATTTACGG	CTCCGACCGC	GCGGGCTACA	ACTACCTGTT
	351	GCGGCGGGCG	AGCGCATCG	CCACATACCA	AATCAACGGC	ATCCCGGTGT
	401	CGACGCGCT	GGCGGATACG	GGCAATGCCA	ACACCGCGCG	CTATGAGCGG
45	451	GTAGAGATCG	TGCGCGCGCT	GGCGGGGCTG	CGCGACGGCA	CGGGGAGCGC
	501	TTCTGCCACC	GTCARCTCTG	TACGCAAAAC	CCGACCGCGC	AAGCCATTGT
	551	TGGAATCGCG	CCCGCAAGCC	GCGACCGCA	AACATTTCCG	CTCGGGCGCG
	601	GACCTATCGG	GCAGCTGAA	CGCCGAGCG	ACGCTGCGCG	CGCGCGTGT
	651	TTCCACCTTC	GACGCGCGCG	ACTCGTGGCG	CGAGCTGAA	CCGACGCGCG
	701	ATGCCGAACT	CTACGGCATT	TTGGAATACG	ACATCGCACG	GCAAAACCGC
50	751	GTCCACGACG	GCATGACTCA	CCAGCAGCGG	AAGAAGAACG	CAGACGCGCC
	801	GCTCAGCTAC	GCGGTGTAG	ACAGCCAGAG	TTATGCCACC	GCCTTCGGCC
	851	CAAAAGACAA	CCCGGCCACA	AATTGTCGA	ACAGCCGCAA	CCGTGCGCTC
	901	AACCTGTTTC	CGGGCATAGA	ACACCGCTTC	AACCAAGACT	GGAACATCAA
	951	AGCGCAATAC	GACTACACCC	GTAGCGGCTT	CGCGCAGGCC	TACGGTGTGG
55	1001	CAGCGGTACT	TTCCATCGAC	CACAGCAGTC	CGCGCACCGA	CCTGATTCCC
	1051	GGTTATTGGC	AGCGcgatcc	GCGCACCCAC	AGCGCCAGCA	TGTCAITGAC
	1101	CGGCAATAC	CgcttGTTCG	GCGCGAGACA	CGATTTAATC	GCGGGATACA
	1151	ACGGCTACAA	ATACGCGCAG	AACAATAAGC	GCGAAGCGAG	CATCATTCOC
	1201	AACGCCATTTC	CCAAAGCGCTA	CGAATTTTTC	CGCACGGGGC	CCTATCCGCA
60	1251	GCCATCATCG	TTTGCCCAAA	CCATCCGCGA	ATACGACACC	AGGCGGCARA
	1301	TCTGCGGCTA	TCTCGCCACC	CGTTTCGCGC	CGCGCGACA	CCTTTCGCTG
	1351	ATATCGCGCG	GCAGATACAG	CGCTACCGC	GCAGCGAGCT	ACACAGCGCG

1401	CACACAAGGC	ATGACCTATG	TGTCGCCAA	CGGTTTACC	CCCTACACAG
1451	GCATCGTGT	CGATCTGAC	GGCAACCTGT	CGCTTTACGG	CTCGTACAGC
1501	AGCCTGTTCG	TCCCGCAATT	GCAAAAGAC	GAACACGGCA	GCTACCTGTG
1551	ACCCGTAACC	GGCAACAATC	TGGAAAGCGA	CATCAAGAGC	GAATGGCTTA
1601	AAGGCGGTCT	GAAACGATCC	GCCGCCGTGT	ACCGCGCCCG	TAAACAAAC
1651	CTCGCCACCG	CAGCAGGACG	CGACCAGAGC	GGCAACACT	ACTATCGCGC
1701	CGCCAAACCA	GCCAAACACC	ACGGCTGGGA	AATCGAAGTC	GGCGGCCGCA
1751	TCAAGCCCGA	ATGGCAGATA	CAGGACAGCT	ACAGCCAAAG	CAAACCCCGC
1801	GACCACAGAC	GCAGCCGCTC	GAAACCCGAC	AGGCTACCG	AACGACGCTT
1851	CAAACTCTTC	ACCGCTTACC	ACTTAGCCGC	CGAAGCCCC	AGCGGCCGGA
1901	CCATCggTGC	GGGTGTGCGC	CGGCAGGGCG	AAACCCACAC	CGACCCAGCC
1951	GGGCTCGCA	TCCCAACACC	CGCCGCCAAA	CGCCGCCCGC	TCCGCAACAG
2001	CGCCGAGAA	GCTACGCCG	TGCGCCGACAT	CATGGCGCGT	TACCGCTTCA
2051	ATCCGCGCAC	CGACCTGCG	CTGAACGCTG	ACACCTGTT	CAACAAACAC
2101	TACCGACACC	AGCCCGACCG	CGACAGCTAC	GGCGCACTGC	GGACAGTGAA
2151	CGCGCGCTTT	ACCTATCGGT	TAAATAA		

This corresponds to the amino acid sequence <SEQ ID 672; ORF23ng-1>:

1	MTRFKYSLLF	AALLPVYAQA	DVSVSDDPKP	QESTELPTIT	VTADRASSN
51	DGYTVSGTHT	PFGLEMTLRE	IPQSVSVITS	QOMRDQNIKT	LDRALLQATG
101	TSRQIYGS DR	AGYNYLFARG	SR IANYQING	IPVADALADT	GNANTAAYER
151	VEVVRGVAGL	PDGTGEPSAT	VNLVRKHPT R	KPLFEVRAEA	GNRKHFGLGA
201	DVSGSLNAEG	TLRGRLVSTF	GRGDSWRQLE	RSRDAELYGI	LEYDIAPQTR
251	VHAGMDYQQA	KETADAPLSY	AVYDSQGYAT	AFGPKDNPAT	NWSNSRNRAL
301	NLFAGIEHRF	NQDWKLKAEY	DYTRSFRFPQ	YGVAGVLSID	HSTAATDLIP
351	GYWHADPRTH	SASMSLTGKY	RLFGREHDLI	AGINGYKYAS	NKYGERSIIP
401	NAIPNAYEFS	RTGAYPQPS	FAQTIPOYDT	RRQIGYGLAT	RFRADNLISL
451	ILGGRYSRYR	AGSYNSRTQG	MTLYSARNFT	PYTGI VFDTL	GNLSLYGSYS
501	SLEVPQLQKD	EHGSYLKPTV	GNNLEADIKG	EWLEGLRNAS	AAVYRARKNN
551	LATAAGRDQS	GNTYYRAANQ	AKTHGWEIEV	GGRTIPEWQI	QAGYSQSKPR
601	DDQGSRLNPD	SVYPERSEKFL	TAYHLAPEAP	SGRTIGAGSVR	ROGETHTDPA
651	ALAIPIEPAK	ARAVANSRQK	AYAVADINAR	YRNPTELS	LNVNLENKHX
701	YRTQDRHSY	GALRTVNAAF	TYRFK*		

ORF23ng-1 and ORF23-1 show 95.9% identity in 725 aa overlap:

35	orf23-1.pep	10	20	30	40	50	60
	MTRFKYSLLFAALLPVYAQA	DVSVSDDPKPQESTELPTIT	VTADRASSN	DGYTVSGTHT			
	orf23ng-1	10	20	30	40	50	60
	MTRFKYSLLFAALLPVYAQA	DVSVSDDPKPQESTELPTIT	VTADRASSN	DGYTVSGTHT			
40	orf23-1.pep	70	80	90	100	110	120
	PLGLPMTLREIPQSVSVITS	QOMRDQNIKT	LDRALLQATG	TSRQIYGS	DRAGYNL	FARG	
	orf23ng-1	70	80	90	100	110	120
	PLGLPMTLREIPQSVSVITS	QOMRDQNIKT	LDRALLQATG	TSRQIYGS	DRAGYNL	FARG	
45	orf23-1.pep	130	140	150	160	170	180
	SR IANYQINGIPVADALADT	GNANTAAYER	VEVVRGVAGL	PDGTGEPSAT	VNLVRKRLTR		
	orf23ng-1	130	140	150	160	170	180
	SR IANYQINGIPVADALADT	GNANTAAYER	VEVVRGVAGL	PDGTGEPSAT	VNLVRKRLTR		
50	orf23-1.pep	190	200	210	220	230	240
	KPLFEVRAEAGNRKHFG	LADVSGSLNTEG	TLRGLVST	FGRGDSWR	QLERSRDA	EYLYGI	
	orf23ng-1	190	200	210	220	230	240
	KPLFEVRAEAGNRKHFG	LADVSGSLNTEG	TLRGLVST	FGRGDSWR	QLERSRDA	EYLYGI	
55	orf23-1.pep	250	260	270	280	290	300
	LEYDIAPOTRVHAGMDYQ	QAKETADAPLSY	AVYDSQGYAT	AFGPKDNPAT	NWSNSRNRAL		
	orf23ng-1	250	260	270	280	290	300
	LEYDIAPOTRVHAGMDYQ	QAKETADAPLSY	AVYDSQGYAT	AFGPKDNPAT	NWSNSRNRAL		
60	orf23-1.pep	310	320	330	340	350	360
	NLFAGIEHRFNQDWKLKAEY	DYTRSFRFPQYGVAGVLSI	DHSTAATDLIP	GYWHADPRTH			
	orf23ng-1	310	320	330	340	350	360
	NLFAGIEHRFNQDWKLKAEY	DYTRSFRFPQYGVAGVLSI	DHSTAATDLIP	GYWHADPRTH			

-385-

		310	320	330	340	350	360
		370	380	390	400	410	420
5	orf23-1.pep	SASVSLIGKYRLFGREHDLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYPOPAS					
	orf23ng-1	SASMSLTGKYRLFGREHDLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYPOPSS					
		370	380	390	400	410	420
		430	440	450	460	470	480
10	orf23-1.pep	FAQTIPQYGTTRRQIGGYLATRFRAADNLSLILGGRYTRTGYSDSRFQMGTYVSANRFT					
	orf23ng-1	FAQTIPQYDTRRQIGGYLATRFRAADNLSLILGGRYSRYRAGSYNSRTQMGTYVSANRFT					
		430	440	450	460	470	480
		490	500	510	520	530	540
15	orf23-1.pep	PYTGTIVFDLTGNLSLYGSYSLSLVFPVQSQKDEHGSYLPVTTGNLLEAGIKGEWLEGRINAS					
	orf23ng-1	PYTGTIVFDLTGNLSLYGSYSLSLVFPVQSQKDEHGSYLPVTTGNLLEADIKGEWLEGRINAS					
		490	500	510	520	530	540
20		550	560	570	580	590	600
	orf23-1.pep	AAVYRARKNNLATAAGRDPSGNTTYRAAHQAKTHGWEIEVGGRITEPEWQIQAGYSQSKTR					
	orf23ng-1	AAVYRARKNNLATAAGRDQSGNTTYRAAHQAKTHGWEIEVGGRITEPEWQIQAGYSQSKFR					
25		550	560	570	580	590	600
		610	620	630	640	650	660
	orf23-1.pep	DQDGSRLNPDSPVPSRFFKLTAYHFAEAPSGWTIGAGVVRWQSETHTPDPAALRIIPNPAAK					
30	orf23ng-1	DQDGSRLNPDSPVPSRFFKLTAYHFAEAPSGRTIGAGVVRWQSETHTPDPAALRIIPNPAAK					
		610	620	630	640	650	660
		670	680	690	700	710	720
	orf23-1.pep	ARAADNSRQKAYAVADIMARYRFNPRAEISLNVNLFNKHYRTQPDHRSYGALRTVNAAF					
35	orf23ng-1	ARAVANSRQKAYAVADIMARYRFNPRTELSLNVNLFNKHYRTQPDHRSYGALRTVNAAF					
		670	680	690	700	710	720
40	orf23-1.pep	TYRFXK					
	orf23ng-1	TYRFXK					

In addition, ORF23ng-1 shows significant homology with an OMP from *E.coli*:

45	sp P16869 FHUE_ECOLI OUTER-MEMBRANE RECEPTOR FOR FE(III)-COPROGEN, FE(III)-FERRIOXAMINE B AND FE(III)-RHODOTRULIC ACID PRECURSOR >gi 1651542 gnl FID d1015403 (D90745) Outer membrane protein FhuE precursor [Escherichia coli]
	>gi 1651545 gnl FID d1015405 (D90746) Outer membrane protein FhuE precursor [Escherichia coli] >gi 1787344 (AE000210) outer-membrane receptor for Fe(III)-coprogen, Fe(III)-ferrioxamine B and Fe(III)-rhodotrulic acid precursor [Escherichia coli] Length = 729
50	Score = 332 bits (843), Expect = 3e-90 Identities = 228/717 (31%), Positives = 350/717 (48%), Gaps = 60/717 (8%)
	Query: 38 TITVTADRTASSN--DGYTVSGTHTPFLGPLMTRLIPQSVSVITSSQMRDQNIKTLDRLAL 95
	T+V TA + +Y+V+T + MT R+IPQSV+++ Q+M DQ ++TL +
55	Sbjct: 43 TVIVEGATAPDDGENDYSVTSTAGTKMQTQDIPQSVTVIVSQMRDEDDQLQLTGEVM 102
	Query: 96 LQATGTSRQIYSGSDRAGYNILFARGSRRIANYQINGIP-----VADALADTGNANTAA 147
	G S+ SDR Y ++RG +I NY ++GIP + DAL+D A
60	Sbjct: 103 ENTLLGSKSQADSDRALY---YSRGFIQIDNYMVDGIPTYFESRWNLGDLALSDM-----AL 154
	Query: 148 YERVEVVRGAVGLPDGTGEPSTNLVRKHPTKPLF-EVRAEAGNRKHFGLGADVSGSL 206
	+ERVEVVRG GL GTC PSA +N+VRKH T + +V AE G+ Ad+ L
65	Sbjct: 155 FERVEVVRGATGLMTGTGNPSAANNVRKHATSRFEGKDVSAEYGSWNKERYVADLQSLP 214
	Query: 207 NAEGTILGRVLSTFGRGDSWRQLERSDAELYGILEYDIPATQTRVHAGMGOYQAKETADA 266
	+G +R R+V + DSW S GI++ D+ T + AG +YO+ +
	Sbjct: 215 TEDGKIRARIVGGYQNLDNSWLDRYNSEKTFSGIVDADLGLDITLLSAGYEQRIDVNSPT 274
70	Query: 267 PLSYAVYDSQGYATAFGPKDNPATNWSNRRNALNLFAGIEHRFNQDWKLKAEYDYTRSR 326

```

      +++ G + ++      + A +W+ +      +F ++ +F W+      ++
Sbjct: 275 WGLPLRNNITDGSNSYDRARSTAPDWAYNDKEINKVFMTLKQQFADTQWATLNATHSEVE 334
5  Query: 327 F--RQPHYGVAGVLSIDHSTAA--TDLIPGY-----WHADPRTHSA-SMSLTGKYRLFG 374
      F + Y A V D      ++ PG+      W++ R A +      G Y LEF
Sbjct: 335 FDSKMMYVDAYVNKADGMLVGPYSNYGPGFDYVGGTGWNSGKRKVDALDLFDAGSYELFG 394
10 Query: 375 REHDLIAGINGKYKASNKYGER--SIIPNAIPNAYEFSRTGAYPQSPSEFAQTIPQYDTR 432
      R+H+L+ G Y +N+Y      +I P+ I + Y E+ G +PQ      Q++ Q DT
Sbjct: 395 RQHNLNMG-GSYSKQNNRYFSSWANIFPDEIGSFYNFN--GNFPQTDWSPQSLAQDDTH 451
15 Query: 433 QIGGYLATREAADNLSLILGGYRSYRAGSYNSRTQGMTY-VSANRFTPYTGIVFDFXXX 491
      Y ATR AD L LILG RY+ R +      +TY + N TEY G+VFD
Sbjct: 452 MKSLYAATRVTLADPLHLILGARYTNWRVDT-----LTYSMKKNHTTYPAGLVFDIND 504
20 Query: 492 XXXXXXXXXXKXEPOLQKDEHGSYLKPVGTGNLEADIKGEWLEGRNLNSAAVYRARNNL 551
      F PQ +D G YL P+TGNN E +K +W+ RL + A++R ++N+
Sbjct: 505 NWSTYASYSIFQPQNDRSSGKYLAFTGNNYELGLKSDWMNRLTTTALFRIEQDNV 564
25 Query: 552 ATAAGR---DQSGNTYYRAANQARKTHGWEIEVGGRTPEWQIQAGYSQSKPRDQDGSRLN 608
      A + G      +G T Y+A +      + G E E+ G IT WQ+ G ++      D +G+ +N
Sbjct: 565 AQSTGTPPGSNGETAYKAVDGTVSGKVEFELNGAITDNWQLTFGATRYIAEDNEGNAV 624
30 Query: 609 PDSVPERSFKLFTAYHLAPEAPSGRTIGAGVRRQGETHTDPAALRIPNPAKARAVANSR 668
      P ++P + K+ET+Y L P P      T+G GV Q +TD      P RA
Sbjct: 625 P-NLPRTTVKMFTSYRL-PVME-LTVGGGVNQNVRVYDTV-----TPYGTFFR-----E 672
Query: 669 OKAYAVADIMARYRNPTELNLNDLNFKNHYRTQPDHR-SYALRITVNAFTYRF 724
      Q +YA+ D+ RY+      L NW+NLE+K Y T +      YG R +      TY+P
Sbjct: 673 CGSYALVDLFTFYQVTKNFSLQGNVNNLFEDKTYDNTVEGSIVYGTFRNFSITGYQF 729

```

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF23-1 (77.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 15A shows the results of affinity purification of the His-fusion protein, and Figure 15B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 15C) and for ELISA (positive result). These experiments confirm that ORF23-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 80

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 673>:

```

1  ATGCGCACGG CAGTGGTTTT GCTGTTGATC ATGCGCATGG CGGCTTCGTC
51  GGCATATGATG CCGGAATGG TGTGCGGCGG CGTGTGCGCC GGAACGGCAA
101 TCATATCCAA CGCAGCCGAA CAACGGCGCG TCATGCGCTC GAGTTTGTCC
45 151 ACGCGGAGCA CGCTGCTTTC GCGCGGCA ATCATACCTT CGTCTTGGGA
201 AACGGGGATA AACCGCGCAC TCAACCGCC GACCGCGCTC GAAGCCATCA
251 TGCCGCGTTC TTTCACGGCA TCGTTCACGA ATGCCAAGC TGCTGTTGTG
301 CCGTGCCTAC CGCAGACGCT CAAGCCCAT TTTTCAGAA TGCGTGCAC
351 TTAGTGCAGC AGGGGG..

```

This corresponds to the amino acid sequence <SEQ ID 674; ORF24>:

```

1  MRTAVVLLI MPMAASSAMM PEMVCAVSP GTAIISKPTE QTAVMASSL
51  SVSTPASAAA IIPSSSETGI NAPLKPPTAL EAIMPPFFTA SFSNAKRAVV
101 PCVPQTLKPI XSRMRATXSP TG..

```

Further work revealed the complete nucleotide sequence <SEQ ID 675>:

```

1   ATGCGCACGG  CAGTGGTTTT  GCTGTTGATC  ATGCCGATGG  CGGCTTCGTC
51  GGCAATGATG  CCGGAAATGG  TGTGCGCGGG  CGTGTGCGCG  GGAACGGCAA
101 TCATATCAAA  CGCGAACGAA  CAAACGGCGG  TCAATGGCTT  GASTTTGTTC
151 AGCGTCAGCA  CGCTTCTTTC  GCGCGCGCGA  ATCATACCTT  CGCTTTGGGA
201 AACGGGATA  AACGCGCCAC  TCACAACCCC  GACCGCGCTG  GAAGCATCAT
251 TGCCTCTTT  TTTACGGGCA  TCGTTACGCA  ATGCCAAGCG  TCGTGTGTGT
301 CGGTGCTTAC  CGCAGACGCT  CAAAGCCATT  TCTTCAAGAA  TCGCTGCCAC
351 TGAGTCGCGG  ACGGCGGGGG  TCGCGCGCAG  CGACAAATCG  AGAATACCAA
401 ACGGGATATT  CAGCATTTTT  GAGGCTTCGC  GGCGCATGAG  TTGCGCCACG
451 CGGGTAATTT  TGAAAGCAGT  TTTCTTCACT  ACTTCCGGA  CTTCGGTCAA
501 TGTGTTGCA  TCTGAATTTT  CCAACGCGCG  TTTTACGACA  CCGTGGCGCG
551 ATACGCGGAC  ATTGATAACG  GCATCCGCTT  CGCCGGAACC  ATGAACGCGG
601 CCGCGCATAA  ACGGGTTGTC  TTCCACCGCG  TTGCAGAAC  CGACATTTT
651 AGCGCAGCG  AACCTTCGG  CGGTGATTTC  CGCGTGGCT  TTGACGGTTT
701 CGCCGCGCAG  CTTGACCGCA  TCGATATTGA  TACCGGCAG  CGTACTCGCG
751 ATATTGATGG  AGCTGCACAC  AATATCGGTA  GTCTTCATCG  CTTCGGGAAT
801 GGAGCGGATT  AACCTTCAT  CGNAAGCGGA  CATCCCTTT  TGCACCAACG
851 CGGAAAAACC  GCGGATAAAA  GACACACCGA  TGGCTTTGG  AGCTTTATTC
901 AAGATTGCG  CACGCTGAC  GTAA

```

This corresponds to the amino acid sequence <SEQ ID 676; ORF24-1>:

```

1   MRTAVVLLLI  MPMAASSAMM  PEMVCAGVSP  GTAIISKPT  QTAVMASSLS
51  SVSTPASAAA  IIPSSSETGI  NAPLKPTAL  EAIMPPFT  SFSNAKAADV
101 PCVPOTLKPI  SSRMRATESP  TAGVGASDK  RPNPFI3IF  EASRPMSSPT
151 RVILKAVFET  TSATSVNVVA  SEFSNAAFT  PGPDPTTIL  TASASPEPNA
201 PAINGLSSTA  LQNTTILAQP  KPSGVISAVR  LTVSPASLTA  SILIPARVL
251 IIMELHTISV  VFIAAGMERI  NTSSEGDIP  CTNAEKPIK  DTPMALAALS
301 KVCATILT*

```

Computer analysis of this amino acid sequence gave the following results:

30 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF24 shows 96.4% identity over a 307 aa overlap with an ORF (ORF24a) from strain A of *N.*

meningitidis:

```

35  orf24a.pep  10      20      30      40      50      60
      MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISKPT QTAVMASSLS NVSTPASAAA
      |||||
35  orf24      10      20      30      40      50      60
      MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISKPT QTAVMASSLS NVSTPASAAA
      |||||

40  orf24a.pep  70      80      90      100     110     120
      IIPSSSXTGINAPLKPTALEAIMPPFFTASFNSAKAAVPCVPQTLKPISSSRMRATESP
      |||||
40  orf24      70      80      90      100     110     120
      IIPSSSXTGINAPLKPTALEAIMPPFFTASFNSAKAAVPCVPQTLKPISSSRMRATESP
      |||||

45  orf24a.pep  130     140     150     160     170     180
      TAGVGASDKSRIPNGIFSI FEASRPMSSPTRVILKAVFETTSATSVNVVASEFSNAAFT
      |||||
45  orf24      130     140     150     160     170     180
      TAGVGASDKSRIPNGIFSI FEASRPMSSPTRVILKAVFETTSATSVNVVASEFSNAAFT
      |||||

50  orf24a.pep  190     200     210     220     230     240
      PGPDPTTLITASASPEPKNAPAI XGLS XALQNTTILAQKPKSSVISXVRLMVS PASLTA
      |||||
50  orf24      190     200     210     220     230     240
      PGPDPTTLITASASPEPKNAPAI XGLS XALQNTTILAQKPKSGVISAVRLTVS PASLTA
      |||||

55  orf24a.pep  250     260     270     280     290     300
      SILIPARVL IIMELHTISV VFIAAGMERI NTSSEGDIP CTSAEKPIK DTPMALAALS
      |||||
55  orf24      250     260     270     280     290     300
      SILIPARVL IIMELHTISV VFIAAGMERI NTSSEGDIP CTNAEKPIK DTPMALAALS
      |||||

```

```

orf24a.pep  KVCATLTX
            |||||
orf24       KVCATLTX

```

5 The complete length ORF24a nucleotide sequence <SEQ ID 677> is:

```

1  ATGCGCACGG  CAGTGGTTT  GCTGTTGATC  ATGCGGATGG  CGGCTTCGTC
51  GGCAATATAG  CCGGAAATGG  TGTGCGCGGG  TGTGTCGCGG  GGAACGCGAA
101  TCATATCCAA  NCCGACGGAA  CAAACGCGGG  TCATCGCTTC  GAGT'TATCC
151  AACCTCAGCA  CGCTCGCTTC  GGGCGGCGCA  ATCATACCTT  CGTCTCGGA
201  NACGGGGGTA  AACGCGCCAC  TCAAACGCGC  AACCGGCGTC  GAAGCATCA
251  TGCGCGCCTT  TTTCAACGCA  TCGTTCAGCA  ATGCCAAGCG  TGCTGTGTGT
301  CGGTGGGTAC  CGCAGACGCT  CAAACCCATT  TCCTCAAGAA  TGGCGGCCAC
351  CGAGTGGCGG  ACGGCAAGGG  TCGGTGCCAG  CGACAGATCG  AGAATACCAA
401  ACGCGATATT  CAGCATTTT  GAGCTTCGCG  GCGCGATGAG  TTGCGCCACG
451  CGGCTAATTT  TGAAGCGGGT  TTCTTTCACA  ACTTCGGCAA  CTTGCGTCAA
501  TGTGTTGCA  TCGCAATTT  CCAAGCGGCG  TTTTAAGACA  CCGCGCGCGG
551  ATAACGCGAC  ATTAATCACA  GCATCCGCTT  CGCCTGAGCC  GTGAAACGCG
601  CCGGCCATAN  ACGGTTGTCT  TTCNCCGCGG  TTGCGAAGCA  CGACGATTTT
651  GCGCGAGCGG  AAACCTTCTA  GTGTGATTTT  ANCCGTGGGT  TTGATGTTTT
701  CGCGCGCAG  TCTGACGCGG  TCCATATTGA  TACCGGCGCG  CGTACTGCGG
751  ATATTGATGG  AGCTGCACAC  GATATCAGTA  GTCTTCATCG  CTTGCGGAAT
801  GGAACGGATN  AACACCTCGT  CAGAAGCGCA  CATACCTTTT  TGCACCAAGG
851  CGGAAAGGCC  GCCAATAAAA  GACACGCCGA  TGGCTTTGGC  AGCCTTATCC
901  AAAGTTTGGC  CCAACGTGAC  GTAA

```

25 This encodes a protein having amino acid sequence <SEQ ID 678>:

```

1  MRTAVVLLLI  MPMAASSAMM  PEMVCAGVSP  GTAIISXPT  QTAVIASSLS
51  NVSTPASAAA  IIPSSSXTGI  NAPLKPPTAL  EAIMPPFFTA  SFSNAKAUV
101  PCVPQTLKPI  SSRMRATESP  TAGVGASDKS  RIFNGIFSIF  EASRPMSSPT
151  RVILKAVFFT  TSATSVNVVA  SEFSNAAPFT  PGDPPTFLIT  ASASPEP*NA
201  PAIKGLSSXA  LQNTTLAQF  KPSSVISXVR  LMVSPASLTA  SILIPARVLF
251  ILMELHTISV  VFIASGMERX  NTSSEGDIPF  CTSAEKPPFK  DTPMALAALS
301  KVCATLT*

```

It should be noted that this protein includes a stop codon at position 198.

ORF24a and ORF24-1 show 96.4% identity in 307 aa overlap:

```

35  orf24a.pep  10 20 30 40 50 60
      MRTAVVLLLIIMPMMAASSAMMPMVCGVSPGTALISXPTQTAVIASSLSNVSTPASAAA
      orf24-1  10 20 30 40 50 60
      MRTAVVLLLIIMPMMAASSAMMPMVCGVSPGTALISXPTQTAVIASSLSNVSTPASAAA

40  orf24a.pep  70 80 90 100 110 120
      IIPSSSXTGINAPLKPPTALEAIMPFFFTASFSNAKAUVPCVPQTLKPISSRMRATESP
      orf24-1  70 80 90 100 110 120
      IIPSSSXTGINAPLKPPTALEAIMPFFFTASFSNAKAUVPCVPQTLKPISSRMRATESP

45  orf24a.pep  130 140 150 160 170 180
      TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNAAPFT
      orf24-1  130 140 150 160 170 180
      TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNAAPFT

50  orf24a.pep  190 200 210 220 230 240
      PGDPPTLITASASPEPKNAPAIKGLSSKALQNTTILAQPKPSSVISXVRIMVSPASLTA
      orf24-1  190 200 210 220 230 240
      PGDPPTLITASASPEPKNAPAIKGLSSKALQNTTILAQPKPSSVISXVRIMVSPASLTA

55  orf24a.pep  250 260 270 280 290 300
      SILIPARVLPILMELHTISVVFIASGMERXNTSSRGDIPFCTSAEKPPFKDTPMALAALS
      orf24-1  250 260 270 280 290 300
      SILIPARVLPILMELHTISVVFIASGMERINTSSRGDIPFCTSAEKPPFKDTPMALAALS

```


5 orf24a.pep KVCATLTX
| | | | |
orf24-1 KVCATLTX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF24 shows 96.7% identity over a 121 aa overlap with a predicted ORF (ORF24ng) from *N.gonorrhoeae*:

10 orf24.pep MRTAVVLLLLIMPMAASSAMPMEMVCAGVSPGTAIISKPTQETAVMASSLSVSTPASAAA 60
orf24ng MRTAVVLLLLIMPMAASSAMPMEMVCAGVSPGTAIMSKPTEQAVMASSLSVNTPASAAA 60
15 orf24.pep IIPSSSETGINAPLKPPTALEAIMPFFFTASFSNAKAAVVCVPQTLKPISRRMRATXSP 120
orf24ng IIPSSSETGINAPLKPPTALEAIMPFFFTASFSNAKAAVVCVPQTLKPISRRMRATESP 120
orf24.pep TG 122
| :
20 orf24ng TAGVGASDKSRMPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVRLTASEFSSAALT 180

The complete length ORF24ng nucleotide sequence <SEQ ID 679> is:

1 ATGGGCACGG CGGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
51 GGCAGATGATG CCGGAAATGG GTGTGCGGGG CGTGTGCGCG GGACACGCCAA
25 101 TCATGTCCAA ACCAACGGAG CAGACGGCGG TCATGGCTTC GAGTTTGTCC
151 AGCGTCAACA CGCTGCGTC GCGGCGGCGA ATCATACTT GTCTTCGGA
201 AACGGGGTA AACGCGCGC TCACACGCG GACCGCGTG GAGACCATCA
251 TGC CGCCTT TTTACGGCA TCGTTACAG ATGCCAAGC TCGCTTGTG
301 CCGTGCCTAC CGCAGAOCCT CAAGCCCAT TCTTCAAGAA TCGCGCCAC
351 CGAGTCGCGC ACGCGCGGGG TCGGTGCCAG CGACAATCG AGAATCGCGA
30 401 ACGGATATT CAGCATTTT GAGGCTTCG GACCGATGAG TTCGCCACG
451 CCGGTGATT TGAAGCGGT TTTCTTCACT ACTTCGGGA CCTCGGTGAC
501 GCTGACCGCG TCGGAATTT CAGCGCGCG TTTGACCAG CCTGGACCGG
551 ATACGCGGAC ATTAATCACA GCATCCGCTT CGCCGAGCC GTGGAACGCA
601 CCGGCCATAA ACGGATTGTC TTCCACGCG TTGCAGAAC CGACGATTTT
35 651 GCGCGACCGG AACCTTCGG GTGTGATTT AGCGTGCTT TTGATGTTT
701 CGCCTGCCAG CTTGACGCGA TCCATATTGA TACCGGACG CGTGCTCCG
751 ATATTGATGG AGCTGCACAC GATATCGGTA GTTTTCATCG CTTGCGGAAC
801 GGAACGGATC AACACCTCAT CGAAGGCGA CATACTTTT TGACACAGCG
851 CGGAAAGGCC GCGGATAAAG GACAGCGCGA TGGCTTTGGC TGCTTGTGC
40 901 AAGCTCTGCG CCACGCTGAC ATAA

This encodes a protein having amino acid sequence <SEQ ID 680>:

1 MRTAVVLLLLIMPMAASSAMPMEMVCAGVSP GTAIMSKPTE QAVMASSLS
51 SVNTPASAAA IIPSSSETGI NAPLKPPTAL EAIMPPFTA SFSNAKAAVV
45 101 PCVPQTLKPI SSRMRATESP TAGVGASDKS RMPNGIFSIF EASRPMSSPT
151 RVILKAVFFT TSATSVRLTA SEFSSAALT PGDPPTLP ASASPEPWIA
201 PAINGLSSSTA LQNTTILAQP KPSSGVISAVR LMVSPASLTA SLIPARVLP
251 ILMELHTISV VFIASTGTERI NTSSEGDIPP CTSAEKPPIK DTFMALAALS
301 KVCATLT*

ORF24ng and ORF24-1 show 96.1% identity in 307 aa overlap:

50 orf24-1.pep 10 20 30 40 50 60
MRTAVVLLLLIMPMAASSAMPMEMVCAGVSPGTAIISKPTQETAVMASSLSVSTPASAAA
| | | | |
orf24ng MRTAVVLLLLIMPMAASSAMPMEMVCAGVSPGTAIMSKPTEQAVMASSLSVNTPASAAA
10 20 30 40 50 60
55 orf24-1.pep 70 80 90 100 110 120
IIPSSSETGINAPLKPPTALEAIMPFFFTASFSNAKAAVVCVPQTLKPISRRMRATESP
| | | | |
orf24ng IIPSSSETGINAPLKPPTALEAIMPFFFTASFSNAKAAVVCVPQTLKPISRRMRATESP
70 80 90 100 110 120
60

-390-

		130	140	150	160	170	180
orf24-1.pep	TAGVGASDKSRIPNGIFSI	FEASRPMSSPTRVILKAVF	FTTSATSVNVVASEF	SNAAF	TT		
5	orf24ng	TAGVGASDKSRMPNGIFSI	FEASRPMSSPTRVILKAVF	FTTSATSVNRATSEFSSA	LT		
		130	140	150	160	170	180
		190	200	210	220	230	240
10	orf24-1.pep	PGPDTPTLITASASPEPXNAPAINGLS	STALQNTTILAQPKPSGVISAVRITV	SPASLTA			
	orf24ng	PGPDTPTLITASASPEPXNAPAINGLS	STALQNTTILAQPKPSGVISAVRIMV	SPASLTA			
		190	200	210	220	230	240
		250	260	270	280	290	300
15	orf24-1.pep	SILIPARVLEIIMELHTISVVF	FIASGMERINTSS	EGDIPFCTNAEKPP	IKDTPMALAALS		
	orf24ng	SILIPARVLEIIMELHTISVVF	FIASGTERINTSS	EGDIPFCTSAEKPP	IKDTPMALAALS		
		250	260	270	280	290	300
20	orf24-1.pep	KVCATLTX					
	orf24ng	KVCATLTX					

Based on this analysis, including the presence of a putative leader sequence (first 18 aa – double-
underlined) and putative transmembrane domains (single-underlined) in the gonococcal protein,
it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could
be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 81

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 681>:

1	..ACCGACGTGC	AAAAAGATT	GGTCGGCGAA	CAACGCAAGT	GGCGCGAGGA
51	AAAAATCAGC	AACTGCCGAC	AAGCCGCGCG	GCAGGCAGAC	CGGCAGGAAT
101	ACGCGAATA	CCTCAAGCTG	CAATGCGACA	CGCGGATGAC	GCGCGAACGG
151	ATACAGTATC	TTGCGGCTA	TTCCATCGAT	TAG	

This corresponds to the amino acid sequence <SEQ ID 682; ORF25>:

1	..TDVQKELVGE	QRKWAQEKIS	NCRQAAAQAD	RQEYAEYLKL	QCQDPMTRER
51	IQYLRGYSID	*			

Further work revealed the complete nucleotide sequence <SEQ ID 683>:

1	ATGTATCGGA	AACCTATTGC	GCTGCCGTTT	GGCCTGCTTG	TTCGCCGTTG
51	CGGCAGGGAA	GAACCGGCCA	AGGCATTGGA	ATGCGCCAAAC	CCGCGCGTGT
101	TGCAAGGCAT	ACGCGGCAAT	ATTCAAGAAA	CGCTCAGCGA	GGAGACGGGT
151	CTTTTTCGCG	GCGAAGACGG	CAGGCAAGTT	GTGATGCGCG	ACAAAATTAT
201	CGCCCGCGCG	TACGTTTTCG	CGTTTTCGTT	GGAAACAGCT	TGCGAAAACG
251	AGGAAGCGCG	GCGCAAGTTC	TGTATCGCG	ATTGGAACAT	TACCGTGCGG
301	TCTGAAACGC	TTGCGGATGC	CAAGGCAAA	AGCCCGCTGT	TGTACGGGGA
351	AACTGCTTTG	TGCGGATATTG	TGCGGCGAGAA	GACGGCGCGG	AATGTCGAGT
401	TTAAAGACGG	CGTATTGACG	CGAGCGCTCC	CGTCTCTGCG	CGTCAALAGC
451	GGTCAGACGG	CATTGTGCGA	CAACACGCTC	GGTATGCGCG	CGCAACGCTC
501	GTCTGCGCGG	CTGCTGCCTT	ACGGCGTGAA	GAGCATCGTG	ATGATAGACG
551	GCAAGGCGGT	GAARAAAGAA	GACGCGGTCA	GGATTTTGAG	CGGAAAAGCC
601	CGTGAAGGAG	AACCGTCCAA	ACCCACGCGC	GAAGACATTT	TGGAACACAA
651	TGCGCGCGGC	GCGGATGCGG	GCGTACCCCA	AGCCGACAGAA	GGCGCGCCCG
701	AACCGGAAAT	CCTGCATCCT	GACGACGCGG	AGCGTGCCGA	TACCGTTTAC
751	GTATCAAGGG	GCGAAGTGGG	AGAGGCGCGC	CTACAAACAT	AGCGTGCGGA
801	ATCGGAATTT	ACCAAACTTT	GCGGAGGACT	CGATACCGAC	GTGCAAAAAG
851	AGTTGCTCGG	CGAACAAACG	AACTGCGCGC	AGGAAAAAAT	CAGCAACTCG
901	CGACAAGCGG	CGCGCGAGCG	AGACCGGCGG	GAATACGCGG	AATACCTCAA
951	GCTGCAATGC	GACACGGGGA	TGACGCGCGA	ACGGATACAG	TATCTTCGCG
1001	GCTATTCCAT	CGATTAG			

This corresponds to the amino acid sequence <SEQ ID 684; ORF25-1>:

```

1 MYRKLIALPF ALLLAAAGRE EPPKALECAN PAVLOGIRGN IQETLTQEAR
51 SFAREDGROF VDADKIIAAA YGLAFSLAHE SETQEGGRTF CIADINITVP
101 SETLADAKAN SPILYGETAL SDIVRGKPTG NVEKIDGVLV AAVRILPVKD
151 GQAFVQNVV GMAQTLASA LLPYGVKSIV MIDGKAVKKE DAVRILSGKA
201 REEPEPKPTP EDILEHNAAG GDAGVQAAE GAPEPEILHF DDGERADTVT
251 VSRGEVEEAR VQNRQAESEI TKLWGLDLD VQKELVGEQR KWAQEKISNC
301 RQAAQADRQ EYAEYLKLCQ DTRMTREIRIQLRGSYID*

```

Computer analysis of this amino acid sequence gave the following results:

10 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF25 shows 98.3% identity over a 60aa overlap with an ORF (ORF25a) from strain A of *N.*

meningitidis:

```

15 orf25.pep TDVQKELVGEQRKWAQEKISNCRQAAQAD
orf25a VTVSRGEVEEARVQNRQAESEITKLWGLDLDVQKELVGEQRKWAQEKISNCRQAAQAD
250 260 270 280 290 300
20 orf25.pep RCEYAEYLKLCQDTRMTREIRIQLRGSYIDX
orf25a RCEYAEYLKLCQDTRMTREIRIQLRGSYIDX
310 320 330

```

The complete length ORF25a nucleotide sequence <SEQ ID 685> is:

```

25 1 ATGTATCGGA AACTCATTGC GCTCGCGTT GCCCTGCTGC TTGCGCGCTTG
51 CGGCAGGGAA GAAACGCCCA AGGCATTGGA ATGCGCCAC CCOCGCTGT
101 TGCANGCAT ACGCNGCAAT ATTCCAGGAA CGCTCACGCA GGAAGCGCGT
151 TCTTTTCGCGC GCGAAGACNG CANGCAGTTT GTCGATCGCG ACNAAATTAT
201 CGCGCGCGCC TANGNTNNGN NGNTNTCTTT GGAACAACCT TCGGAAACCG
251 AGGAAGCGGG GCGCACGTTT TGTNTCGCGC ATTTGAACAT TACCGTGCGG
301 TCTGAAROGC TTGCCGATGC CAAGGCAAA AGCCCCCTGC TGTAGCGGGA
351 AACCGCTTTG TGGGATATTG TCGCGGCAAA GACGGCGGCG AATGTCGAGT
401 TTAAGACCGG CATTATGACG GCACGCGTCC GGTATCGCGG CGCAACCGCT
451 GGTCAAGAGG CATTTGTCGA CAACAGCGTC GGTATCGCGG CGCAACCGCT
501 CTCTCGCGCG TTGCTGCGCTT ACGCGGTGAA GAGCATCGTG ATGATGACGC
551 CCAAGCGCGT AAAAAAGAA GACCGGTCA GGATTMTGAG CNGANAAGCC
601 CGTGAANAAG AACCGTCCAA ANCCNNGCCC GAAGACATT TGGACATAAA
651 TGCGCGCGGA GGGGATGCGA ACGTACCCCA AGCGGAGGAA GACGCGCGCG
701 AACCGGAAAT CCGTATCTCT GACGACGGCG ACGGTGCGGA TACCGTTACC
40 751 GTATCACGGG GCGAAGTGGA AGAGGCGGCG GTACAAAACC ACGCTGCGGA
801 ATCCGAAATT ACCAACTTTT GGGGAGGACT CGATACGCGC GTGCAAAAAG
851 AGTGTGTCGG CGAANAACGC AAGTGGCGCG AGGAAAAAAT CAGCACTGCA
901 CGACAAGCGC CGCGCGAGCG AGACCGGCG GATACGCGG AATACCTCAA
951 GCTGCAATGC GACACGCGGA TGACGCGGGA ACGGATACAG TATCTTCGCG
45 1001 GCTATTCCAT CGATTAG

```

This encodes a protein having amino acid sequence <SEQ ID 686>:

```

1 MYRKLIALPF ALLLAAAGRE EPPKALECAN PAVIQXIRXN IQETLTQEAR
51 SFAREDXKOF VDADKIIAAA XXXXXSLEHA SETQEGGRTF CXADINITVP
101 SETLADAKAN SPILYGETAL SDIVRGKPTG NVEKIDGVLV AAVRILPVKD
151 GQAFVQNVV GMAQTLASA LLPYGVKSIV MIDGKAVKKE DAVRILSGKA
201 REEPEPKPTP EDILEHNAAG GDAGVQAAE GAPEPEILHF DDGERADTVT
251 VSRGEVEEAR VQNRQAESEI TKLWGLDLD VQKELVGEQR KWAQEKISNC
301 RQAAQADRQ EYAEYLKLCQ DTRMTREIRIQLRGSYID*

```

ORF25a and ORF25-1 show 93.5% identity in 338 aa overlap:

```

55 orf25a.pep MYRKLIALPFALLLAAAGRE EPPKALECANPAVLQXIRXNIQETLTQEARSFAREDXKOF
orf25-1 MYRKLIALPFALLLAAAGRE EPPKALECANPAVLQIRXNIQETLTQEARSFAREDXKOF
10 20 30 40 50 60

```

[illegible]

ORF25 shows 100% identity over a 60aa overlap with a predicted ORF (ORF25ng) from

35 *N.gonorrhoeae:*

	orf25 pep		TDVOKELVGGEQRKWAQEKISNCRQAAQAAD	308
	orf25 ng		VTVSRGVEEEARVNQRAASEITKLWGGSLDTDVOKELVGGEQRKWAQEKISNCRQAAQAAD	308

40

	orf25 pep		RQEYAEYLKLCQCDTMTTRERIRCYLGRYSID	60
	orf25 ng		RQEYAEYLKLCQCDTMTTRERIRCYLGRYSID	338

The complete length ORF25ng nucleotide sequence <SEO ID 687> is:

45	51	ATGTAATGGGA	AACATCTATTC	GCTGCGGCTTT	GCCTCTGCTGC	TTGCAAGCGTG
	51	CGCAGGAGGAA	AGACCGCGCCA	AGGCGT'TGGA	ATTGCGCCAGAA	CGCCCGGCTGT
	101	TGCGAGCAAT	ACCGCGCAGT	ATTGAGGAAA	GCTCTCAGCA	GGAGAGCGGCT
	115	TC'TT'TCGGCG	AAGCGGATCG	CAGGCAAGCTT	TCGTGTCGCG	ACAAAT'TTAT
	201	CGCGCGCCGC	TACGGT'TTGG	GCT'TTCTTCT	GGAAACACGCT	TGCGAAACGCG
50	251	AGGAAAGCGG	CGCGACGCTC	TGTATCGCTTC	ATT'TGGAACAT	TACGTCGCCG
	301	TCTGAACCGC	TGTCGCGATG	CGGCGCAAA	AGGCCCTCTG	TGTAT'TGGGGA
	351	ATCAGCT'TTTC	CGACAGCATG	TGCAGCAAGAC	GACGGGCGCG	AATGTCGAGT
	401	T'AAAGAACGG	CGTAT'TGACG	CGACGCGTCC	GCTTCTCGTG	CGCCAAAGACT
	451	CGTCGGAGCG	CATTAT'ATCG	CAACACGCTG	GGTAT'TGGCGA	CGCCAAAGCA
55	501	G'TCTCGCCGC	TGTC'TGCTTT	ACGCGCTGAT	GAGATCTCGT	ATGATGACGAG
	551	CGAAGCGCGT	GA'CAAAAGAA	CGACGGTCTA	GGTTT'TTGG	CGCCAAAGCC
	601	CGTGAAGARG	AACCGCT'CAAA	ACCCACGCCCC	GAGRCAT'TTT	TGGAACACAA
	651	TGCGCGCGCG	CGCGATGCGG	CGTGATCCCCA	AGCGCGAGAA	GGCGCACCCG
	701	NACCGGAAAT	CCCTGATCCG	CGACGCGTCC	AGTTCGCGCA	TACCGT'TACC
60	751	G'TATCGCAAT	CGCGAAGACG	CGCGGAGGAT	AGTTCGCGCA	AGTTCGCGCA
	801	ATCCGGAATC	ACCA'AAATC	CGCGGAGCAT	CGATACCCAG	TGCGAA'AAAT
	851	AGT'TGG'CGG	CGACGACGCG	AAGTGGGAGC	AGGAAA'AAAT	CGGAaactCG
	901	cgACAAGCGC	CGCGCGCAGC	AGACGCGGAC	GAAATCCCTAA	AATATCTCTAA
	951	GCTTCAACAT	GA'ACGCGGGA	TGACGCGCGA	ACgga'TACAG	TATCTTTCGGC
1001	GCTAT'TTCAT	CGATTAG				

65 This encodes a protein having amino acid sequence <SEQ ID 688>:

```

1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQDIRGS IQETLTQEAR
51 SFAREDGROF VDADKIIAAA YGLAFSLEHA SETQEGGRFT CIADLNITVP
101 SETLADAEAN SPFLYGETSL ADIVQOKTGG NVEFKDGVLT AAVRFLPAKD
151 ARTAFIDNTV GMATQTLASAA LLPYGVKSTV MIDGKAVTKE DAVRVLSGKA
201 REEPFSKPTP EDILEHNAAG GDAGVQAAE GAPEPELLHP DDVERADTWT
251 VSRGEVEEAR VQNRASEI TKLWGGLD TDVQKELVGEOR KWAQEKISNC
301 RQAAAQADRO EYAEYLKLCQ DTRMTREIRIQ YLRGYSID*

```

ORF25ng and ORF25-1 show 95.9% identity in 338 aa overlap:

```

10 orf25-1.pep      10      20      30      40      50      60
      MYRKLIALPFALLLAACGREPPKALECANPAVLQDIRGNIQETLTQEARSFAREDGROF
      |||||||||||||||||||||||||||||||||||||||||||||
orf25ng      10      20      30      40      50      60
      MYRKLIALPFALLLAACGREPPKALECANPAVLQDIRGSIQETLTQEARSFAREDGROF
      |||||||||||||||||||||||||||||||||||||||||||||

15 orf25-1.pep      70      80      90      100     110     120
      VDADKIIAAAYGLAFSLEHASSETQEGGRFTFCIADLNITVPSETLADAEANSPFLYGETAL
      |||||||||||||||||||||||||||||||||||||||||||||
orf25ng      70      80      90      100     110     120
      VDADKIIAAAYGLAFSLEHASSETQEGGRFTFCIADLNITVPSETLADAEANSPFLYGETSL
      |||||||||||||||||||||||||||||||||||||||||||||

20 orf25-1.pep      130     140     150     160     170     180
      SDIVRQKTGGNVEFKDGVLTAAVRFLPVKDGQTAFVDNTVGNAAQTLAALLPYGVKSIV
      :|||:|||||||||||||||||||||||||||||||||||||
orf25ng      130     140     150     160     170     180
      ADIVQOKTGGNVEFKDGVLTAAVRFLPAKDARTAFINTVGNATQTLAALLPYGVKSIV
      |||:|||||||||||||||||||||||||||||||||||||

25 orf25-1.pep      190     200     210     220     230     240
      MIDGKAVKEDAVRILSGKAREEPPSKPTPEDILEHNAAGGDAGVQAAEGAPEPELLHP
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||
orf25ng      190     200     210     220     230     240
      MIDGKAVKEDAVRILSGKAREEPPSKPTPEDILEHNAAGGDAGVQAAEGAPEPELLHP
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||

30 orf25-1.pep      250     260     270     280     290     300
      DDGERADTWTVSRGEVEEARVQNRASEITKLWGGLD TDVQKELVGEOR KWAQEKISNC
      || |||||:|||||:|||||:|||||:|||||:|||||:|||||
orf25ng      250     260     270     280     290     300
      DDVERADTWTVSRGEVEEARVQNRASEITKLWGGLD TDVQKELVGEOR KWAQEKISNC
      || |||||:|||||:|||||:|||||:|||||:|||||:|||||

40 orf25-1.pep      310     320     330     339
      RQAAAQADROEYAEYLKLCQDTRMTREIRIQYLRGYSIDX
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||
orf25ng      310     320     330
      RQAAAQADROEYAEYLKLCQDTRMTREIRIQYLRGYSIDX
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||

```

Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF25-1 (37kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 16A shows the results of affinity purification of the GST-fusion protein, and Figure 16B shows the results of expression of the His-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 16C), ELISA (positive result), and FACS analysis (Figure 16D). These experiments confirm that ORF25-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 16E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF25-1.

Example 82

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 689>

```

5      1  ATGCAGCTGA  TCGACTATTC  ACATTCAATT  TTCTCGGTTG  TGCCACCCCTT
      51  TTGGCACTG  GCACCTGCGG  TCATTACCG  CCGCGTACTG  CTGCTTTTAG
     101  GCATCGGTAT  TCTGGTGGCG  GTTGCCCTTT  TGGTGGCGCG  CAACCCCGTC
     151  GACGGCTTGA  CACACCTGAA  AGACATGGTC  GTGGCTTGG  CTTGGTCAGA
     201  CGsYgATTGG  TCGCTGGGCA  AACCAAAAT  CTTGGTTTTT  CkGATACTTT
     251  TGGGTATTTT  TACTTCCCTG  CTGACCTACT  CCGGCAGCAA  T.....

10      //

     851  .....AC  TTGCTGGTA
     901  TTGGGCGGCA  CTTGCGGCGT  CTTTGGCGST  GTTCTCTGCA  CGCTCGGCAC
     951  GATTAAACCC  GCGGACTATC  CCAAAGCCGT  TTGGCAGGGT  GCGAAATCTA
    1001  TGTTCGGCGC  AATCGCCATT  TTAATCCTCG  CTTGGCTCAT  CAGTACGGTT
    1051  GTGGGGGAAA  TGCACACCGG  CGATTACCTG  TCCACACTGG  TTGGCGGCTA
    1101  CATCGATCCC  GCGTTCCCTG  CGGTATCCCT  CTCTCTGCTC  GCCACGGTGA
    1151  TGGCGTTTTC  CACAGGCACA  AGCTGGGGGA  CGTTGGCAT  TATGCTGCGG
    1201  ATTGCGCGCG  CCATGGCGGT  CAAAGTCGAA  CCGCGGCTGA  TTATCCCGTG
    1251  TATGTCGCGC  GTAATGGCGG  GGGCGGTATG  CGGCGACCAT  TGCTCGCCCA
    1301  TTTCCGACAC  GACCATCTCG  TCGTCCACGG  GCGCGGGCTG  CAACCACATC
    1351  GACCACGTTA  CTTGCGCACT  GCTTTAGCGC  TTAACCGTTG  CGCGCGCGCG
    1401  CGCATCGGGC  TACCTCGCAT  TGGGTCTGAC  AAAATCCGCG  CTGTTGGGCT
    1451  TTGGCACGAC  AGGCATTGTA  TTGGCGGTGC  TGATTTTCT  GTTGAAGAT
    1501  AAAAAA..

```

25 This corresponds to the amino acid sequence <SEQ ID 690; ORF26>:

```

30      1  MQLIDYSHSF  FSVVPFFLAL  ALAVITRRLV  LSLGIGILXX  VAFLVGGNEV
      51  DGLTHLKDMV  VGLAWSDXDW  SLGPKILFV  XILLGIPTSL  LTYGSSN...

     251  .....TSLV
     301  FGGTCGVFAV  VLCTLTGKT  ADYPKAVWQ  AKSMRGAI  LILAWLSTV
     351  VGEHHTGDYL  STLVAENIHP  GFLVLEILL  ASVMFATGT  SMGTGIMLP
     401  IAAAMAVKVE  PALIIICMIS  VMAGAVGSD  CSPISDTTIL  SSTGARCNH
     451  DRYTSQLEPYA  LTVAAAAAG  YLALGLTKS  LLGFSTTIV  LAVLFLLLD
     501  KK..

```

35 Further work revealed the complete nucleotide sequence <SEQ ID 691>:

```

40      1  ATGCAGCTGA  TCGACTATTC  ACATTCAATT  TTCTCGGTTG  TGCCACCCCTT
      51  TTGGCACTG  GCACCTGCGG  TCATTACCG  CCGCGTACTG  CTGCTTTTAG
     101  GCATCGGTAT  TCTGGTGGCG  GTTGCCCTTT  TGGTGGCGCG  CAACCCCGTC
     151  GACGGCTTGA  CACACCTGAA  AGACATGGTC  GTGGCTTGG  CTTGGTCAGA
     201  CGGCGATTGG  TCGCTGGGCA  AACCAAAAT  CTTGGTTTTT  CTGATACTTT
     251  TGGGTATTTT  TACTTCCCTG  CTGACCTACT  CCGGCAGCAA  TCAGGCGTTT
     301  GCGGACTGGG  CAAAACGGCA  CATTAAAC  CCGCGGGCG  CGAAAATGCT
     351  GACCGCTGCT  CTGCTGTTG  TAACCTTTAT  CGACGACTAT  TTCCACAGTC
     401  TCGCGCTCGG  TCGGATTGCC  CGCCCGGTGA  CCGACAGATT  TAAAGTTTCC
     451  CGCACCAAAC  TCGCCTACAT  CCTCGACTCC  ACTGCGCTC  CTATGTGGT
     501  GCTGATGCC  GTTTCAGCT  GGGGCGGCT  GATTATCGC  ACGCTTGGG
     551  GACTGCTCGT  TACCTACAAA  ATCACCGAAT  ACACGCGAT  GGGGCGTTT
     601  GTGCGCATGA  GCTTGATGAA  CTAATTAGCA  CTGTTGGCC  TGATTTGGT
     651  GTTGTGCTG  CGATGTTTT  CTTGACAT  CGGCTGAGT  GCACGTTGG
     701  AACAGCGC  GTTACAGCA  GCCAGATC  AACGCGCT  TTCAAGCTG
     751  ACCAAGGTC  GTGTTAGCG  ACTGATTA  CCGSTTTGG  CCTTAATCG
     801  CTCAACGGTT  TCGCGCATGA  TCTACACGG  CGCGCAGGA  AGCGAAACCT
     851  TCAGCATTTT  GGGGCACTT  GAAACACGG  ACGTAACAC  TTGCTGGTA
     901  TTGCGCGGCA  CTTGCGGCGT  CTTTGGCGST  GTTCTCTGCA  CGCTCGGCAC
     951  GATTAAACCC  GCGGACTATC  CCAAAGCCGT  TTGGCAGGGT  GCGAAATCTA
    1001  TGTTCGGCGC  AATCGCCATT  TTAATCCTCG  CTTGGCTCAT  CAGTACGGTT
    1051  GTGCGCGAAA  TGCACACCGG  CGATTACCT  TCCACACTGG  TTGCGGGGCA
    1101  CATCCATCCC  GCGTTCTGCG  CGGTATCCT  CTCTCTGCTC  GCCACGCTGA
    1151  TGGCGTTTTC  CACAGGCACA  AGCTGGGGGA  CGTTGGCAT  TATGCTGCGG
    1201  ATTGCGCGCG  CCATGGCGGT  CAAAGTCGAA  CCGCGGCTGA  TTATCCCGTG
    1251  TATGTCGCGC  GTAATGGCGG  GGGCGGTATG  CGGCGACCAT  TGCTCGCCCA
    1301  TTTCCGACAC  GACCATCTCG  TCGTCCACGG  GCGCGGCTG  CAACCACATC

```

-395-

```

1351 GACCACGTTA CCTCGCAACT GCCTTACGCC TTAACCGTTG CCGCCGCCGC
1401 CGCATCGGGC TACCTCGCAT TGGGTCTGAC AAAATCCGGC CTGTGGGGCT
1451 TTGGCAGCAG AGGCATTGTA TTGGCGGTGC TGATTTTCT GTTGAAGAT
1501 AAAAACGCG CCAACGCTG A

```

5 This corresponds to the amino acid sequence <SEQ ID 692; ORF26-1>:

```

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRLV LSLGIGILVG VAFVLGGNPFV
51 DGLTHLKMV VGLAWSGDW SLGPKILVF LILGIFTSL LTYSGSNQAF
101 ADWAKRHIK RRGAKMLTAC LVFTFIDDY FHSLAVGAIA RPTDKFKVS
151 RTKLAYILDS TAAPMCVLP VSSMGASIIA TLAGLLVYTK ITEYTPMGTF
201 VAMSLMYYA LFLALIMFVV AWFSEFDIGSM ARFEQAALNE AHDSTAVSDA
251 TKGRVYALII PVLLALIASTV SAMIYTTGAQA SETFSILGAF ENTDVNTSLV
301 FGSTCGVLAV VLCTLGTIKT ADYPKAVWQG AKSMFGAIAI LILAWLISTV
351 VGEHMTGDYL STLAVGNIHP GLFVLVILFL ASVMAFATGT SWGTFGIMLP
401 IAAAHAVKVE PALIIPCMSA VMAGAVCGDH CSPISDITIL SSTGARCNIH
151 451 DHVTSQLPYA LTVAAAAAG YLALGLTKSA LLGFGTGTGIV LAVLIFLLKD
501 KKRANA*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical transmembrane protein III1586 of *H. influenzae* (accession number P44263)

ORF26 and HI1586 show 53% and 49% amino acid identity in 97 and 221 aa overlap at the

20 N-terminus and C-terminus, respectively:

```

Orf26 1 MQLIDYSHSFFSVVPPFLALALAVITRVRVXXXXXXXXXXVAFVLGGNPVDGLTHLKMV 60
M+LID+ S S +S+VP LA+ LA+ TRRV L +L V
HI1586 14 MELIDFSSSVWSIVFALLAIILAIATIRRVLSLGSIIIGSLMSDWQIGSAFNYLVKNV 73

Orf26 61 VGLAWSDXDWSLGGPKILVFXKILLGIFTSLTYSGSN 97
V L +D + +LG+ T+LL SGN
HI1586 74 VSLVYADGEIN-SNNMIVLFIILLGLVLTALLTVSGSN 109

//

Orf26 86 IFTSLTYSGS--NTSLVFGGTGCVFAVLVCLT--GTIKTADYPKAVWQGAQSMFGXXXX 141
+F+ L T+ + TSLV GG C + L + + +Y++ G KSM G
HI1586 299 VFSVLGTFTENTVVGTSVLVGGFCSIIISTLLIILDRQVSVEYVRSVIVIGIKSMGSAIAI 358

Orf26 142 XXXXXXSTTVGEMHTGDYLSTLVAGNIHPGLFVLVILFLASVMAFATGTSGWTFGIMLP 201
+VGM+M TG YLS+LV+GNI FLFVLVIL+L +MAF+TGTSGWTFGIMLP
HI1586 359 LFFAWTINKIVGDMQTKYLSLVSGNIPMQFLFVLVILGAAMAFSTGTSGWTFGIMLP 418

Orf26 202 IAAAHAVKVEPALIIPCMSAVMAGAVCGDHCSPISTDTILSSTGARCNIHDVTSQXXXX 261
IAAAMA P L++PC+SAVMAGAVCGDHCSPISTDTILSSTGA+CNIHDVTS+Q
HI1586 419 IAAAHAAANAPPELLLPCLSAVMAGAVCGDHCSPISTDTILSSTGARCNIHDVTSQQLPYA 478

Orf26 262 XXXXXXXXXXXXXXXXXXXXKALLGFGTGTGIVLAVLIFLLKDK 302
S L GF T + L V+IF +K +
45 HI1586 479 ATVATATSIGIIVGFTYSGLAGFAATAVSLIVIFAVKKR 519

```

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF26 shows 58.2% identity over a 502aa overlap with an ORF (ORF26a) from strain A of *N. meningitidis*:

```

10 20 30 40 50 60
orf26.pep MQLIDYSHSFFSVVPPFLALALAVITRRLVLSLGGIGILXXVAFVLGGNPVDGLTHLKMV
orf26a MQLIDYSHSFFSVVPPFLALALAVITRRLVLSLGGIGILVGVAFVLGGNPVDGLTHLKMV
10 20 30 40 50 60
orf26.pep VGLAWSDXDWSLGGPKILVFXKILLGIFTSLTYSGSNXX-----
orf26a VGLAWSGDWWSLGGPKILVFXKILLGIFTSLTYSGSNQAFADWAKRHIKNRGAKMLTAC
70 80 90 99

```

	orf26.pep	-----
5	orf26a	<u>L</u> VFVT <u>F</u> IDDYFHS <u>L</u> AVGAXARPVTDKFKVSRAKLAIYLDSTAAPMCVLPVSSWSGASIIA
		130 140 150 160 170 180
10	orf26.pep	-----
	orf26a	<u>T</u> IAGLLVLYKITEYTPMGTFVAMSLMNYALFALIMVFVWAFSFDIGSMARFEQAALNE
		190 200 210 220 230 240
15	orf26.pep	-----100 110-----TSLV
	orf26a	AHDETA <u>V</u> SDSGWGRVYAL <u>I</u> PVLLALIASTVSAMIYTTGAQASETFSILGAFENTDVNTSLV
		250 260 270 280 290 300
20	orf26.pep	120 130 140 150 160 170 FGGTCGVFAVVLCTLGITKADYPKAVWQGA <u>S</u> MFGAIALILAWLISTVVGEMHTGDYL
	orf26a	FGGTCGVFAVVLCTLGITKADYPKAVWQGA <u>S</u> MFGAIALILAWLISTVVGEMHTGDYL
		310 320 330 340 350 360
25	orf26.pep	180 190 200 210 220 230 STLVAGNIHPGFLPVILFLLASVMAFATGT <u>S</u> WGTFGIMLPFAAAMAVKVEPALIIPCMSA
	orf26a	STLVAGNIHPGFLPVILFLLASVMAFATGT <u>S</u> WGTFGIMLPFAAAMAVKVEPALIIPCMSA
		370 380 390 400 410 420
30	orf26.pep	240 250 260 270 280 290 VMAGAVCGDHCSPISTDTILSSSTGARNHIDHVT <u>S</u> OLPYALTVAAAAAGYIALGLTKSA
	orf26a	VMAGAVCGDHCSPISTDTILSSSTGARNHIDHVT <u>S</u> OLPYALTVAAAAAGYIALGLTKSA
		430 440 450 460 470 480
35	orf26.pep	300 310 LLGFGTTGIVLAVLILFLKDKK
	orf26a	LLGFGTTGIVLAVLILFLKDKK
		490 500

The complete length ORF26a nucleotide sequence <SEQ ID 693> is:

45	1	ATGCAGCTGA	TGACTATTC	ACATTCACTT	TTCTCGGTG	TGCCACCCCT
	51	TTTGCCACTG	GCACCTGCGG	TCATTACCCG	CCGCGTACTG	CTGCTCTTAG
	101	GCATCGGTAT	TCGTGGTCGG	GTGCGCTTTT	TGGTCGGCGG	CAACCCGCTG
	151	GACGTCGTGA	CACACCTGAA	AGACATGTGG	CTGCGCTTGG	CTTGCTCAGA
	201	CGCGATTGG	TGCTGGGCA	AACCAAANT	CTTGTCTTC	CTGACTCTT
	251	TGGGTATTTT	TACTTCCCTG	CTGACTACT	CCGCGACGAA	TCAGGCTTTT
50	301	CGCGACTGGG	CAAAACGGCA	CATTAAAAAC	CGCGCGCGCG	CGAAATAGCT
	351	GACCGCTGCG	CTCGTGTTOG	TAACCTTTAT	CGACGACTAT	TTCCACAGTC
	401	TGCGCGTGG	TGCGNTTGCC	CGCCCGCTTA	CCGACAAGTT	TAAAGTTTCC
	451	CGCGCCAAAC	TGCGCTACAT	CCTCGACTCC	ACTGCCGCGG	CTATGTGCTG
	501	GCTGATGCC	GTTTCAAGCT	GGGGCGGCTC	GATTATCGCC	ACGCTTGCGG
55	551	GACTGCTCGT	TACCTACAAA	ATCACCNAAT	ACACGCGGAT	GGGACGCTT
	601	GTCGCCATGA	GCCTGATGAA	CTATTACGCA	CTGTTTGCCC	TGATTATGTT
	651	GTTGCTCGTC	GCATGGTTCT	CCTTGACAT	CGGCTCGATG	GCACGTTTGG
	701	AACAAGCCGC	GTTGAACGAA	GCCCCAGATG	AAACTGCGGT	TTACAGACGC
	751	AGCTGGGGCA	GGGTTTAAGC	ATTGATTATT	CCGCTTTTGG	CCTTAATCGC
60	801	CTCAACGGTT	TCCGCCATGA	TCTACACCGG	TGCACAGGCA	AGCGAAACCT
	851	TCAGCAITTT	GGGTGCATTY	GAAATAACGG	ACGTGAACAC	TTGCTGTGTA
	901	TTGCGCGGCA	CTTGCGGCGT	GCTTGCGCTC	GTCCTCTGCA	CGCTGGGACA
	951	GATTAAATC	GCGGATTATC	CCAAAGCGGT	TTGGCAGGGT	CGGAAATCCA
	1001	TTTTCGCGCG	AATCGCCATT	TTAATCCTTG	CTGTGCGCAT	CACTACGGCT
65	1051	GTCGCGGAAA	TGCACACAGG	CGACTACCTC	TCCACGCTGG	TTGCGGGACA
	1101	CATCCAATCC	GGCTTCTTGN	CCGTCACTCT	TTTCTGCTCT	GCCACGGTGA
	1151	TGGCGTPTGC	CACAGGCACA	AGCTGGGGGA	CGTTGCGCAT	CATGCTGCGG
	1201	ATTGCGCGCG	CCATGGCGGT	CAAAGTCGAT	CCCTCACTGA	TTATCCGCTG
	1251	TATGTCGCGC	GATGATGGCG	GGGCGGTATG	CGGCGACCAC	TGCTCGCCCA
70	1301	TTTCGACAC	GACCATCTCT	TGCTCCACCG	GCGCGCGCTG	CAACCACTAT

1351 GACCACGTTA CNTCGCAACT GCCTTACGCC TTAACCGTTS CCGCGCGCGC
 1401 CGCATCGGNN TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGTT
 1451 TTGGCAGNAC AGGCATTGTA TTGGCGGTGC TGATTTTTCT GTTGAAGAT
 1501 AAAAAACGCG CCAACGCCTG A

5 This encodes a protein having amino acid sequence <SEQ ID 694>:

1 MOLIDYSHSF FSVVPPFLAL ALAVITRRLV LSLGIGILVG VAFVLGGNPV
 51 DGLTHLKMV VGLAWSGDW SLGKPKXLVF LILLGIFTSL LTYSGSNQAF
 101 ADWAKRHIKN RGAKMLTAC LVFVTFIDY FHSLAVGAXA RPTDKFKVS
 151 RAKLAYILDS TAAAPMCVMP VSSWGASIIA TLAGLLVITYK IETEYTPMGT
 201 VAMSLMNYA LFAIMVFPV AMFSFDIGSM ARFEQAALNE AHDETAUSDG
 251 SWGRVYALII PVLALIASTV SAMIYTAQAQ SETPSILGAF ENTDNVNTSIV
 301 FGTCGVLAV VLCTLTGTIKI ADYPKAVWQ AKSMFGAIAL LILAWLISTV
 351 VGEHMTGIDL STLWAGNIHP GFLXVILFLA ASMAFATGT SWGTFGIMLP
 401 TAAAMAVKVD PSIIIPCMSA VMAGVCGDHC CSPISDTTIL SSTGARNCHI
 451 DHVTSOLPYA LTVAAAAASG YLALGLTKSA LLGFGXTGIV LAVLIFLLKD
 501 KKRANA*

ORF26a and ORF26-1 show 97.8% identity in 506 aa overlap:

		10	20	30	40	50	60
20	orf26a.pap	MOLIDYSHSF	FSVVPFLAL	ALAVITRRLV	LSLGIGILVG	VAFVLGGNPV	DGLTHLKMV
	orf26-1	MOLIDYSHSF	FSVVPFLAL	ALAVITRRLV	LSLGIGILVG	VAFVLGGNPV	DGLTHLKMV
		10	20	30	40	50	60
25	orf26a.pap	70	80	90	100	110	120
	orf26-1	VGLAWSGDW	SLGKPKXLVF	LILLGIFTSL	LTYSGSNQAF	ADWAKRHIKN	RGAKMLTAC
		70	80	90	100	110	120
30	orf26a.pap	130	140	150	160	170	180
	orf26-1	LVFVTFIDY	FHSLAVGAXA	RPTDKFKVS	RAKLAYILDS	TAAAPMCVMP	VSSWGASIIA
		130	140	150	160	170	180
35	orf26a.pap	190	200	210	220	230	240
	orf26-1	TLAGLLVITYK	IETEYTPMGT	FVAMSLMNYA	LFAIMVFPV	AMFSFDIGSM	ARFEQAALNE
		190	200	210	220	230	240
40	orf26a.pap	250	260	270	280	290	300
	orf26-1	AHDETAUSDG	SWGRVYALII	PVLALIASTV	SAMIYTAQAQ	SETPSILGAF	ENTDNVNTSIV
		250	260	270	280	290	300
45	orf26a.pap	310	320	330	340	350	360
	orf26-1	FGTCGVLAV	VLCTLTGTIKI	ADYPKAVWQ	AKSMFGAIAL	LILAWLISTV	VGEHMTGIDL
		310	320	330	340	350	360
50	orf26a.pap	370	380	390	400	410	420
	orf26-1	STLWAGNIHP	GFLXVILFLA	ASMAFATGT	SWGTFGIMLP	PIAAMAVKVD	PSIIIPCMSA
		370	380	390	400	410	420
55	orf26a.pap	430	440	450	460	470	480
	orf26-1	VMAGVCGDHC	CSPISDTTIL	SSTGARNCHI	DHVTSOLPYA	LTVAAAAASG	YLALGLTKSA
		430	440	450	460	470	480
60	orf26a.pap	490	500				
	orf26-1	LLGFGXTGIV	LAVLIFLLK	KKRANA*			
		490	500				
65	orf26a.pap						

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRLV LSLGIGILVG VAFVGGNPV
 51 DGLTHLKMV VGLAWADGW SLGKPKILVF LILLGIFTSL LTYSGSNQAF
 101 ADWAKRHKN RCGAKMLTAC LVFVTIDDDY FHSILAVGAI RVPVTKFKVS
 151 RAKLAYILDS TASFMCVIMP VSSWGASIIA TLAGLLVTVK ITEYTPMGFF
 5 VAMSLMNYA LFAIMVFPV AWFSDIGSM ARFEQAALNE AQDETAASDA
 201 TKGRVYALI I PVLALIASTV SAMIYTGAAQ SETFSILGAF ENTDVNTSLV
 301 FGGTCGVLAV VLCTFGTIKT ADYPKAVWQG AKSMFGAIAI LILAWLISTV
 351 VGEHMTGDYL STLAVAGNIHP GLPLVILFLL ASVMAFATGT SWGTFGIMLE
 401 IAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISDITL SSTGARNCHI
 10 DHVTSOLPYA LTVAAAAASG YLALGLTKSA LLGFGTTGIV LAVLIFLLKD
 501 KKRADV*

ORF26ng and ORF26-1 show 98.4% identity in 505 aa overlap:

		10	20	30	40	50	60
15	orf26-1.pep	MQLIDYSHSF	FSVVPFLAL	ALAVITRRLV	LSLGIGILVG	VAFVGGNPV	DGLTHLKMV
	orf26ng	MQLIDYSHSF	FSVVPFLAL	ALAVITRRLV	LSLGIGILVG	VAFVGGNPV	DGLTHLKMV
		10	20	30	40	50	60
20	orf26-1.pep	VGLAWSDGWS	LGKPKILV	FLILLGIFTSL	LLTYSGSNQAF	ADWAKRHKN	RNRGAKMLTAC
	orf26ng	VGLAWADGWS	LGKPKILV	FLILLGIFTSL	LLTYSGSNQAF	ADWAKRHKN	RNRGAKMLTAC
		70	80	90	100	110	120
25	orf26-1.pep	LVFVTIDDDY	FHSILAVGAI	ARVPVTKFKVS	RTKLAYILD	STAAFMV	CVMPVSSWGASIIA
	orf26ng	LVFVTIDDDY	FHSILAVGAI	ARVPVTKFKVS	RAKLAYILD	STAAFMV	CVMPVSSWGASIIA
		130	140	150	160	170	180
30	orf26-1.pep	TLAGLLVTVK	ITEYTPMGTF	VAMSLMNYA	LFAIMVFPV	AWFSDIGSM	ARFEQAALNE
	orf26ng	TLAGLLVTVK	ITEYTPMGTF	VAMSLMNYA	LFAIMVFPV	AWFSDIGSM	ARFEQAALNE
		190	200	210	220	230	240
35	orf26-1.pep	AHDETAASDA	TKGRVYALI	IPVLALIAST	VSAMIYTGAAQ	SETFSILGAF	ENTDVNTSLV
	orf26ng	AQDETAASDA	TKGRVYALI	IPVLALIAST	VSAMIYTGAAQ	SETFSILGAF	ENTDVNTSLV
		250	260	270	280	290	300
40	orf26-1.pep	FGGTCGVLAV	VLCTFGTIKT	ADYPKAVWQG	AKSMFGAIAI	LILAWLISTV	VGEHMTGDYL
	orf26ng	FGGTCGVLAV	VLCTFGTIKT	ADYPKAVWQG	AKSMFGAIAI	LILAWLISTV	VGEHMTGDYL
		310	320	330	340	350	360
45	orf26-1.pep	STLVAGNIHP	GLPLVILFLL	ASVMAFATGT	CSWGTFGIMLE	PIAAAMAVKVE	PALIPCMSA
	orf26ng	STLVAGNIHP	GLPLVILFLL	ASVMAFATGT	CSWGTFGIMLE	PIAAAMAVKVE	PALIPCMSA
		370	380	390	400	410	420
50	orf26-1.pep	VMAGAVCGDH	CSPISDITL	SSTGARNCHI	DHVTSOLPYA	LTVAAAAASG	YLALGLTKSA
	orf26ng	VMAGAVCGDH	CSPISDITL	SSTGARNCHI	DHVTSOLPYA	LTVAAAAASG	YLALGLTKSA
		430	440	450	460	470	480
55	orf26-1.pep	LLGFGTTGIV	LAVLIFLLK	DKKRRANAX			
	orf26ng	LLGFGTTGIV	LAVLIFLLK	DKKRRADVX			
		490	500				
60	orf26-1.pep	LLGFGTTGIV	LAVLIFLLK	DKKRRANAX			
	orf26ng	LLGFGTTGIV	LAVLIFLLK	DKKRRADVX			
		490	500				
65	orf26-1.pep	LLGFGTTGIV	LAVLIFLLK	DKKRRANAX			
	orf26ng	LLGFGTTGIV	LAVLIFLLK	DKKRRADVX			
		490	500				

In addition, ORF26 ng shows significant homology to a hypothetical *H. influenzae* protein:

-400-

```

sp|P44263|YF86_HAEIN  HYPOTHETICAL  PROTEIN  HI1586  >gi|1074850|pir|I|C64037
hypothetical
protein HI1586 - Haemophilus influenzae (strain Rd KW20) >gi|1574427 (U32832) H.
influenzae predicted coding region HI1586 [Haemophilus influenzae] Length = 519
5   Score = 538 bits (1370), Expect = e-152
   Identities = 280/507 (55%), Positives = 346/507 (68%), Gaps = 7/507 (1%)

Query: 1   MQLIDYSHSFFSVVPPFLLALALAVITRRXXXXXXXXXXXXAFLVGGNFPVDGLTHLKMV 60
10  Sbjct: 14 MELIDFSSSVWSIVFALLAIIAIAIRRVLSLSAGIIIGSIMLSMDQIGSAFNYLVKNV 73

Query: 61   VGLAWADGSDWSLGGPKILVFLILLGIFTLLTSGSNGAFADWAKRIKNCRCGAXMLTAC 120
15  Sbjct: 74 VSLVYADGEIN-SNMNIVLFLLLGLVLTALLTVSGSNRAFAEWAQSRIGKRRGAKLLAAS 132

Query: 121  LVFVTFIDDFHSLAVGAIARPVTDKFKVSRAKLAYILDSTASPMCVMPVSSWGSIIA 180
20  Sbjct: 133 LVFVTFIDDFHSLAVGAIARPVTDKFKVSRAKLAYILDSTAAPMCMVMPVSSWGSIIIT 192

Query: 181  TLAGLLVTYKITEYTPMGTFVAMSLMNYALFALIMVFFVWAFSFDIGSMARFEGAAALNE 240
25  Sbjct: 193 LIGGLLATYSITEYTPIGAFVAMSSMNFYAFISIMVFFVAFSFDIASMVRHEKLAKN 252

Query: 241  AQDETAASDATKGRVYALIIIPVLIASIVSAMIIYTGAAQ----SETFSILGAFENTDVN 296
30  Sbjct: 253 EDQLEETGTGKQVRNLIPLVLIATVSMIIYTGAEALADGKVSFVLGTFTENTVVG 312

Query: 297  TSLVFGGTCGVL--AVVLCTFGTIKTADYPKAVMQGAKSMFGXXXXXXXSTVVGEM 354
35  Sbjct: 313 TSLVGGGFCIIISTLLIILDRQVSVEFVRSWIGIKSHGAIILFFAWTINKIVGDM 372

Query: 355  HTGDYLSPLVAGNIHPGLFVLIFLLASVMAFATGTSWGTGFGIMLPAAAMVKEPALI 414
40  Sbjct: 373 QTKRYLSSLVSGNIPMQFLPVILFVLGAAMAFSTGTSWGTGFGIMLPAAAMANAPEL 432

Query: 415  IFCMSAVMAGAVCGDHCSPISDTTLISSTGACRNHIDHVTXXXKXXXXXXXXXXXXX 474
45  Sbjct: 433 LFLCSAVMAGAVCGDHCSFVSDTTLISSTGAKCNHIDHVTQLPYAATVATATSIGYIVV 492

Query: 475  XXXKSALLGFGTTGIVLAVLIFLLKDK 501
50  Sbjct: 493 GFTYSGLAGFAATAVSLIVIFAVKKR 519

```

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*,
 45 and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 83

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 697>:

```

1   ..AAGCAATGGT ATGCCGACGN .AGTATCAAG ACGGAATGG TTATGGTCAA
51  CGATGAGCCT GCCAAATTC TGACTTGGGA TGAAAGCGGC CGATTACTCT
50  101 CGGAAGTCTC TATCGCCAC CATCAACGCA ACGGGTGGT TTTGAGTGG
151  TATGAAGATG GTTCTAAAA GAGCGAAGT. GTTTATCAGG ATGACAGTT
201  GTTCAGGAAA ACCCAGTGGG ATAAGATGG TTATTATATC GAACCCCTGA

```

This corresponds to the amino acid sequence <SEQ ID 698; ORF27>:

```

1   ..KQWYADXSIIK TEMVMVNDEP AKILTWDSEG RLLSELSIRH HORNGVLEW
55  51  YEDGSKKSEK VYQDDKLVKR TQWDKDGYLE EP*

```

Further work revealed the complete nucleotide sequence <SEQ ID 699>:

```

1   ATGAAAAAT TATCTCGGAT TGTATTTTCA ACTGTCCTGT TGGGTTTTTC
51  GGCCGCTTTG CCGGCGCAGA CCTATTCTGT TTATTTTAAAT CAGAACGGAA
101  AGCTGACGGC GACGATGCTC TCTGCCGCTT ATATCAGGCA ATATAGTGTG
60  151 GTGGCGGTTA TTGCGCAGC GCAGGATTTT TATTATCCGT CGATGAAGAA

```

5

10

15

20

25

30

35

40

45

50

55

10 20 30 40 50 60
orf27a.pep MKILSRIVFSTVLLGFSAALPAQXYSVYFNQNGKLTATXSSAAYIROYSVAEGIAHQXF

-402-

	orf27-1	 MKKLSRIVFSTVLLGFSAALPAQTYSVYFNQNGKLTATMSAAIYRQYSVVAGIAHAQDF 10 20 30 40 50 60
5	orf27a.pep	70 80 90 100 110 120 XYPSMKYSEPYIVASTQIKSFVPTLQNGMLILWHFNGQKKNMAGGFSKGPEDGEVWVNWYP
	orf27-1	70 80 90 100 110 120 YYPSMKYSEPYIVASTQIKSFVPTLQNGMLILWHFNGQKKNMAGGFSKGPEDGEVWVNWYP
10	orf27a.pep	130 140 150 160 170 180 NGKKSAVMPYKNGLSEGTGRYRNGGKESEIQFQKNKANGVWKQYADGNIKTEMVMVN
	orf27-1	130 140 150 160 170 180 NGKKSAVMPYKNGLSEGTGRYRNGGKESEIQFQKNKANGVWKQYADGSIKTEMVMVN
15	orf27a.pep	190 200 210 220 230 240 DEPAKILTWDESGRLLSLSIRHHQRNGVLEWYEDGSKKXEAVYQDDKILVRKTQWDKDG
	orf27-1	190 200 210 220 230 240 DEPAKILTWDESGRLLSLSIRHHQRNGVLEWYEDGSKKXEAVYQDDKILVRKTQWDKDG
20	orf27a.pep	190 200 210 220 230 240 DEPAKILTWDESGRLLSLSIRHHQRNGVLEWYEDGSKKXEAVYQDDKILVRKTQWDKDG
	orf27-1	190 200 210 220 230 240 DEPAKILTWDESGRLLSLSIRHHQRNGVLEWYEDGSKKXEAVYQDDKILVRKTQWDKDG
25	orf27a.pep	YLIEPX
	orf27-1	YLIEPX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF27 shows 96.3% identity over 82 aa overlap with a predicted ORF (ORF27ng) from

N.gonorrhoeae:

	orf27.pep	KQWYADXSIKTEMVMVNDEPAKILTWDESG	30
	orf27ng	LSEGTGRYRNGGKESEIQFQKNKANGVWKQYADGSIKTEMVMVNDEPAKILTWDESG	193
35	orf27.pep	RLLSLSIRHHQRNGVLEWYEDGSKKSEVYQDDKILVRKTQWDKDGYLEP	82
	orf27ng	RLLSLSIRHHQRNGVLEWYEDGSKKSEVYQDDKILVRKTQWDKDGYLEP	245

The complete length ORF27ng nucleotide sequence <SEQ ID 703> is:

40	1 ATGAAGAAAT TATCTGGAT TGTATTTTCA ATCGTACTGT TGGGTTTTTC
	51 GGCGGCTTTG CGGCGCAGA CCTATTCTGT TTATTTTAAT CAGAACGGGA
	101 AACTGAAGGC GACGATGTCT TCTGCGCTTT ATATCAGGCA ATATAGTGTG
	151 GCGGCGGTA TCGCACGCG CGAGGATTT TATTATCCGT CGATGAAGAA
	201 ATATTCCGAA CTTATATCG TTGCTTCAAC GCAATCAAA TCTTTCTGC
	251 CTACCTTCCA AAACGCTATG TTGATTTTGT GGCATTTTAA TCGTCAGAAA
45	301 AAAATGGCGG GGGCTTTCAG CAGGCTAAG CCGGACGGGG AATGGGTCAA
	351 CTGGTATCCG AACCGTAAAA AATCTCGGT TATGCTTTAT AAAATGGGCT
	401 TGAGTGAGGG TACGGGATAC CGTTATTACC GTAACGGCGG CAAGGAAGC
	451 GAAATCCAGT TTAAGCAAAA TAAGCGAAG CCGCTATGGA AGCAATGGTA
	501 TGCCGATGGA AGTATCAAGA CGGAATGAT TATGGTCAAC GATGAGCCTG
50	551 CCAAAATTCT GACTTGGGAT GAAAGCGGC GATTACTTTT GGAATCTGCT
	601 ATCCGCCACC ATAAACGCAA CGGGTGGTT TTGGAGTGGT ATGAAGATGG
	651 TTCTAAAAAG AGCGAGGCTG TTTATCAGGA TGACAAGTTG GTCAGGAAAA
	701 CCAATGGGA TAAGGATGTT TATTATATCG AACCTGTA

This encodes a protein having amino acid sequence <SEQ ID 704>:

55	1 MKKLSRIVFS IVLLGFSAAL PAQTYSVYFN QNGKLTATMS SAAIYRQYSV
	51 AAGIAHAQDF YYPSMKYSE PYIVASTQIK SFVPTLQNG MLILWHFNGQK
	101 KMAGGFSK GK PDGEVWVNWYP NGKKSAMVPY KNLSEGTGY RYRNGGKES
	151 EIQFQKNK AN GWWKQYADG SIKTEMVMVN DEPAKILTW ESRLLSELS
	201 IRHHRNGV V LEWYEDGSKK SEAVYQDDKI VRKTQWDKG YLIEP*

ORF27ng and ORF27-1 show 98.8% identity in 245 aa overlap:

orf27-1.pep	10 20 30 40 50 60 MKKLSRIVFSTVLLGFSAALPAQTYSVYFNQNGKLTATMSAAIYRQYSVVAGIAHAQDF
-------------	--

	orf27ng	MKKLSRVVFSIVLLGFSAAALPAQTYSVVFNQNGKLTA ^{TMSSAAYITRQY} SVAAGIAHAQDF 10 20 30 40 50 60
5	orf27-1.pep	YYPSMKKYSEPYIVASTQIKSFVPTLQNGMLLWHFNGQKKMAGGSXKGKPDGEWNWNP 70 80 90 100 110 120
	orf27ng	YYPSMKKYSEPYIVASTQIKSFVPTLQNGMLLWHFNGQKKMAGGSXKGKPDGEWNWNP 70 80 90 100 110 120
10	orf27-1.pep	NGKKSAVMPYKNGLSGETGYRYRRNGGKESEIQFKNKANGVKWQYADGSIKTEMVMYN 130 140 150 160 170 180
	orf27ng	NGKKSAVMPYKNGLSGETGYRYRRNGGKESEIQFKNKANGVKWQYADGSIKTEMVMYN 130 140 150 160 170 180
15	orf27-1.pep	DEPAKILTWDESGRLLSEL SIRHHQRNGVLEWYEDGSKSSEAVYDQDKLVKRTQWDKDG 190 200 210 220 230 240
	orf27ng	DEPAKILTWDESGRLLSEL SIRHHKRNGVLEWYEDGSKSSEAVYDQDKLVKRTQWDKDG 190 200 210 220 230 240
20	orf27-1.pep	YLIEPX
	orf27ng	YLIEPX
25	orf27-1.pep	YLIEPX
	orf27ng	YLIEPX

Based on this analysis, including the putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF27-1 (24.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 17A shows the results of affinity purification of the GST-fusion protein, and Figure 17B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA, which gave a positive result, confirming that ORF27-1 is a surface-exposed protein and a useful immunogen.

Example 84

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 705>:

40	51	ATGAAATTTTA	CCAGACACCC	CGT CGGGCA	ATGGGGTTCT	GCCCATTTTA
	101	CTGTCTGGCG	GCTCTGTGAC	GGCGATTGCT	TGGGGTTCTG	
	151	GTACACGGG	AACGCACAG	CTGTCTGGAT	TCTATTGGCA	CGCGCATAGg
	201	ATGATTTTGG	GTTATGCGG	ACTGCTCTTC	ATCGGCTCTC	TCTGACCCG
	251	CTGCGCACT	GACGCGGGC	AGCGCCAC	GGGGGGCGC	GTATCTGGTC
45	301	GGCTTGACTA	CTTTTGTGCT	GGCTGGCGGG	TATCGCGCT	TTATCCGGG
	351	TGGGGGTGCT	TGCGCAACGG	GCATACCTGG	TACGCTGTCT	TTCTGGTAGC
	401	CGCGCTATGG	CTGCTGCTTG	CGCGCTATGG	CGCGCTATGG	
	451	TATGTATGCG	TCTGCGGCT	GTTCGCTCTG	GGGCGACG	ATGGGCGCT
	501	CAACGTCAG	GTGCACAACG	GCAACCTAGG	CGGACTCTTG	AGCGGATTGC
	551	AGTGGCGGCT	GGTGATG			

50 This corresponds to the amino acid sequence <SEQ ID 706; ORF47>:

1 MKFTKHPVWA MAFRPFYSLA ALYGALSVLV WGFGYTGTHX LSGFYWHARE
51 MIWGYAGLVV IAFLLTAVAT WTGQPPTRGG VLVGLTIFWL AARIAAFIPG
101 W GASASGILG TLFFWYGAVC MALPVIRSON QRNYVAVFAL FVLGGTHAAF

151 HVQLHNGNLG GLLSGLQSGLV VM

Further work revealed the complete nucleotide sequence <SEQ ID 707>:

```

1   ATGAAATTTA CCAAGCACCC CGTCTGGGCA ATGGCGTTCG GCCCATTTTA
5   51   TTCGCTGGGG GCTCTGTACG CGCATTTGTC CGTATTGCTG TGGGGTTTCG
101  GCTACACGGG AACGCACGAG CTGTCGGGTG TCTATTGGCA CGCGCATGAG
151  ATGATTTTGGG GTTATGCGCG ACTGGTCGTC ATCGCTCTCC TGCTGACCGC
201  CGTCGCCACT TGGCAGGGGG AGCGGCCACG CGGGGGCGCG GTTCTGGTCG
251  GCTTGACTAT CTTTTGGCTG GCTGCGGCGA TTGCGCGCTT TATCCGCGGT
301  TGGGGTGCCT CGGCAACGCG CATACTCGGT ACGCTGTTTT TCTGCTACGG
10  351  CGCGGTGTGC ATGGCTTTTC CCGTTATCCG TTCGCAGAAT CAACGCAACT
401  ATGTTGCCGT GTTCGCGCTG TCGCTCTTGG GCGGCACGCA TCGCGGCTTC
451  CACGTCACAG TGCACAACGG CAACCTAGGC GGACTCTTGA GCGGATTGCA
501  GTCCGGCTTG GTGATGCTGT CGGGTTTATG CGGTCTGATT GGTACGGCGA
551  TTTATTGCTT TTTTACGTCC AAACGCTTGA ATGTGCGCGA GAITCCCGAT
15  601  CCGAATATGGG TGGCGCAGCG TTGCTGTGGT CTGCCCAPCG TGACTGCCAT
651  GCTGATGGCG CACGGTGTGT TGCTTGGCT GTCTCGCGGT TTTGCGCTTG
701  CGGCAAGTGT GATTTTATCC GTGCAGGTGT ACCGCTGGTG GTATAAACCC
751  GTGTGAAAG AGCGCATGCT GTGGATTCTG TTTGCGCGCT ATCTCTTAC
801  CGGATTGGGG CTGATTGCGG TCGCGCGCTC TTATTTCAAA CCGCGTTCC
20  851  TCAATCTGGG TGTGCATCTG ATCGGGGTG CGGCTATCCG CGTCTGACT
901  TTGGGCATGA TGGCGCGTAC CGCGCTTGGT CATACGGCGA ATCCGATTTA
951  TCCGCCGCCCC AAAGCCGCTC CCGTTCGGTT TTGGCTGATG ATGGCGGCAC
100  1001 CGCGCCGTCCG TATGGTTGCC GTATTTTCTT CCGGCACTGC CTACAGCGAC
1051  AGCATCGCGA CTTCTTGGGT TTTGTTTGCA CTCGCGCTTT TGGTGTATCG
25  1101 GTGGAAGTAT ATTCTTGGC TGATTCTGTC GCCTTCGGAC GGCAGGCCGG
1151 GTTGA

```

This corresponds to the amino acid sequence <SEQ ID 708; ORF47-1>:

```

1   MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WGFYGTGTHE LSGFYWAHAE
51  MINGYAGLVV IAFLLTAVAT WTGPPTTRGG VLVLITFWL AARIAAFIPG
30  101  WGAASAGILG TLFWFYGAVC MALPVIIRSON QRNYYAVFAL FVLGGTHAAF
151  HVQLHNGNLG GLLSGLQSGLV VMVSGEIGLL GTRIISFSTS KRLNVPQIPS
201  PKWVMOASLW LPMILMMLWA MCVLRLMSAV EAAAGVIVET VVVYRWKPV
251  VLKSPMLWL EAGYLTGLG LILVAGASYE PAFLMLAGVIT IGVGGIGVLT
301  LGMWARTALG HTGNPIYEPF KAVEFAFWLM MAATVARMVA VFSSGTAYTH
35  351  SIRTSSVLEA LALLVYAWKY IPWLRPRSD GRG*

```

Computer analysis of this amino acid sequence predicts a leader peptide and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF47 shows 99.4% identity over a 172aa overlap with an ORF (ORF47a) from strain A of *N.*

40 *meningitidis*:

		10	20	30	40	50	60
orf47.pep		MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFYGTGTHXLSGFYWAHAE	MINGYAGLVVIAFLLTAVATWTGPPTTRGGVLVLITFWLAARIAAFIPG				
orf47a		MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFYGTGTHXLSGFYWAHAE	MINGYAGLVVIAFLLTAVATWTGPPTTRGGVLVLITFWLAARIAAFIPG				
		10	20	30	40	50	60
		70	80	90	100	110	120
orf47.pep		IAFLLTAVATWTGPPTTRGGVLVLITFWLAARIAAFIPGWGASASGILGLT	TLFFWYGAVC				
orf47a		IAFLLTAVATWTGPPTTRGGVLVLITFWLAARIAAFIPGWGASASGILGLT	TLFFWYGAVC				
		70	80	90	100	110	120
		130	140	150	160	170	
orf47.pep		MALPVIIRSONQORNYYAVFALFVLGGTHAAFHVOLHNGNLGGLLSGLQSGLV	M				
orf47a		MALPVIIRSONQORNYYAVFALFVLGGTHAAFHVOLHNGNLGGLLSGLQSGLV	M				
		130	140	150	160	170	180
orf47a		GTRIISFSTSKRLNVPQIPSPKWVAQSLWLPMLTAMLAHGVMPWLSAAFAFAGVIT	T				
		190	200	210	220	230	240

The complete length ORF47a nucleotide sequence <SEQ ID 709> is:

```

1 ATGAAATTTA CCAAGCACCC CGTTTGGGCA ATGGCGTTCC GCCCGTTTTA
51 TTCACCTGGG GCTCTGTACG GCGCATTTGC CGTATTGCTG TGGGGTTTCG
101 GCTACACGGG AACGCACGAG CTGTCCGGTT ATGCTTGGCA CGCGCATGAG
151 ATGATTGGGG GTTATCGCGG ACTGGTCTTC TGTCTGACCG GCTCTGTCTG
201 CTGCGCATCT TGGACGGGCG AGCGCGCCAC GCGGGCGCGG GTCTGTCTGC
251 GCTTGACTAT CTTTGTGGCTG GCTGGCGGGA TTGCGGCTT TATCCCGGCT
301 TGGGGTGGCT CGGCAAGCGG CATACTCGGT ACGCTGTTT TCTGTACGGG
351 CGCGGTGTGC ATGGCTTTGC CGGTATACCG TTCGCAGAA TCAACGAAAT
401 ATGTTGCGCT GTTCGCGCTG TCTGCTCTGG GCGGTACGCA CGCGCGGTTC
451 CACGCTCCAG TGCACAAAGG CAACCTAGGC GGACTCTTGA CGGGATTGCA
501 GTCGGGCTTG GTGATGGTGT CGGGCTTTAT CGGTCTGATT GGTACGCGGA
551 TTATTTTCGTT TTTTACGTCC AAACGGTTGA ATGTGCGCGA GATTCCCGAT
601 CCGAAATGGG TGGCGCAGCG TTGCGTGTGG CTGCGCATGC TGACCGCCAT
651 GCTGATGGCG CACGGCGTGA TGCCTTGGCT CTGCGCGGCT TTGCGGTTTG
701 CGCAGGTGCT GATTTTACCG GTGCGAGTGT ACCGCTGGTG GTATAAGCTC
751 GTGTGAAAG AGCGGATGCT GTGGATTCTG TTTGCGGCT ATCTGTTTAC
801 CGGATTGGGG CTGATTGGCG TCGGCGGCTC TTATTTACGA CCGGCTTTCC
851 TCAATCTGGG TGTGCATCTG ATCGGGGTCG GCGGATACGG CGTGCTGACT
901 TTGGGCATGA TGGCGGCTAC CGCGCTCGGT CATACGGGCA ATCCGATTTA
951 TCGCGCGCCC AAAGCGGCTC CGGTGCGCTT TTGCTGATG ATGGCGGCGA
1001 CGCGGCTCG TATGGTTCGC GTATTTCTTC CCGGCATGCG CTACACGAC
1051 AGCATACGCA CTTCTTCGCT TTGTTTGCA CTCGCGCTT TGTGTATGCG
1101 GTGAAGTAT ATTCCTTGGC TGATTGCTCC GCGTCCGAG GGCAGGCCCG
1151 GTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 710>:

```

1 MKFTKHPVWA MAFRPPFSLA ALYGALSVLL WGFYGTGTHE LSGFYWHARE
51 MIWGYAGLVV IAFLLTAVAT WTGQFPTRGG VLVLGTFIWL AARIAAFITG
101 WGAASGILG TLFFWYGAVC MALPVRISQN QRNYVAVFAL FVLGGTHAAF
151 HVQLHNGNLG GLSLGQSGSL VMVSGFIGLI CTRIIISFTS KRLNVQIIFS
201 PKWVAQASLW LFMLTAMLMA HGVMPWLAA FAFAGVITF VQYRWYKPF
251 VLKEPMLWIL FAGYFTGLG LIAVGASYFK PAFLNLGVHL LGVGGIGVLT
301 LGMARTALG HTGNFIYPPF KAVPVAFLWM MAATAVRMVA VFSSGTATYTH
351 SIRTSSVLEA LALLVYAWKY IPWLIRPSD GRFG*

```

ORF47a and ORF47-1 show 99.2% identity in 384 aa overlap:

		10	20	30	40	50	60
or f47a.pap		MKFTKHPVWAMAFRPPFSLAALYGALSVLLWGFYGTGTHELSGFYWHAREMIWGYAGLVV					
or f47-1		MKFTKHPVWAMAFRPPFSLAALYGALSVLLWGFYGTGTHELSGFYWHAREMIWGYAGLVV					
		10	20	30	40	50	60
or f47a.pap		70	80	90	100	110	120
		IAFLLTAVATWTGQFPTRGGVLVLGTFIWLAAARIAAFIPGWGASASGILGTLFFWYGAVC					
or f47-1		IAFLLTAVATWTGQFPTRGGVLVLGTFIWLAAARIAAFIPGWGASASGILGTLFFWYGAVC					
		70	80	90	100	110	120
or f47a.pap		130	140	150	160	170	180
		MALPVRISQNRNYVAVFALFVLGGTHAAHFVQLHNGNLGSLGSLGQSGLMVMSGFIGLI					
or f47-1		MALPVRISQNRNYVAVFALFVLGGTHAAHFVQLHNGNLGSLGSLGQSGLMVMSGFIGLI					
		130	140	150	160	170	180
or f47a.pap		190	200	210	220	230	240
		GTRIIISFTSKRLNVQIIFSPKWVAQASLWLEMLTAMLMAHGVMPWLAAAFAGVITF					
or f47-1		GTRIIISFTSKRLNVQIIFSPKWVAQASLWLEMLTAMLMAHGVMPWLAAAFAGVITF					
		190	200	210	220	230	240
or f47a.pap		250	260	270	280	290	300
		VQYRWYKPFVLKEPMLWILFAGYFTGLGLIAVGASYFKPAFLNLGVHLGVGGIGVLT					
or f47-1		VQYRWYKPFVLKEPMLWILFAGYFTGLGLIAVGASYFKPAFLNLGVHLGVGGIGVLT					
		250	260	270	280	290	300
		310	320	330	340	350	360

	orf47a.pep	LGMARTALGHTGNPIYPKAPVAFWLMMAATAVRMVAVFSSGTAYTHSIRTSVLFA
	orf47-1	LGMARTALGHTGNPIYPKAPVAFWLMMAATAVRMVAVFSSGTAYTHSIRTSVLFA
5		310 320 330 340 350 360
	orf47a.pep	LALLVYAWKYIPWLIRPRSDGRPGX
10	orf47-1	LALLVYAWKYIPWLIRPRSDGRPGX
		370 380

Homology with a predicted ORF from *N.gonorrhoeae*

ORF47 shows 97.1% identity over 172 aa overlap with a predicted ORF (ORF47ng) from *N.gonorrhoeae*:

15	ORF47	MKFTKHPVWMAFRPFYSLAALYGALSVLLWGFGYTGTHELSGFYWHAHEMIWGYAGLVV	60
	ORF47ng	MKFTKHPVWMAFRPFYSLAALYGALSVLLWGFGYTGTHELSGFYWHAHEMIWGYAGLVV	60
20	ORF47	IAFLLTAVATWGGQPTRGGVVLGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC	120
	ORF47ng	IAFLLTAVATWGGQPTRGGVVLGLTAFWLAARIAAFIPGWGAASGILGTLFFWYGAVC	120
	ORF47	MALPVIRSQNRNRYVAVFALFVLGGTHAAHFVQLHNGNLGGLLSGLQSGLV	172
25	ORF47ng	MALPVIRSQNRNRYVAVFALFVLGGTHAAHFVQLHNGNLGGLLSGLQSGLV	180

The ORF47ng nucleotide sequence <SEQ ID 711> is predicted to encode a protein comprising amino acid sequence <SEQ ID 712>:

30	1 MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WFGYTGTHELSGFYWHAHE
	51 MIMGYAGLVV IAFLLTAVAT WGGQPTRGG VLVGLTAFWL AARIAAFIPG
	101 WGAASASGLS TLFFWYGAVC MALPVIRSQNRNRYVAVFALFVLGGTHAA
	151 HVQLHNGNLG GLLSGLQSGV VMVWGFGLI GMKILSPFTS KRLKLPQIPS
	201 PKMVAAHSLW LPLNALIMA HRVMPWLSAA PFPAAGVIFT VQVYAGGITP
	251 IETSCGSAV GICYRLGNSS G

The predicted leader peptide and transmembrane domains are identical (except for an Ile/Ala substitution at residue 87 and an Leu/Ile substitution at position 140) to sequences in the meningococcal protein (see also *Pseudomonas stutzeri* orf396, accession number e246540):

	TM segments in ORF47ng
40	INTEGRAL Likelihood = -5.63 Transmembrane 52 - 68
	INTEGRAL Likelihood = -3.88 Transmembrane 169 - 185
	INTEGRAL Likelihood = -3.08 Transmembrane 82 - 98
	INTEGRAL Likelihood = -1.91 Transmembrane 134 - 150
	INTEGRAL Likelihood = -1.44 Transmembrane 107 - 123
	INTEGRAL Likelihood = -1.38 Transmembrane 227 - 243

Further work revealed the complete gonococcal DNA sequence <SEQ ID 713>:

45	1 ATGAATTTA CCAACATCC CGTCTGGGCA ATGGCGTTCC GCCCGTTTTA
	51 TTCACGTGGG GCACTGTACG GCGCATTTGC CGTATTGCTG TGGGGTTTCG
	101 GCTACACGGG AACGCACGAG CTCTCCGGTT TCTATTGCGA CGCGCATGAG
	151 ATGATTTGGG GTTATGCCG GTCCTGCTGC ATCGCCTTCC TCGTACCGG
	201 CGTCCCACT TGGACGGGAC AGCCGCCAC GAGGCGCGG GTTCTGCTCG
50	251 GCTTACACGC CTTTTGGCTG CCGTCCGCGA TTGCCGCTTT TATCCCGGT
	301 TGGGCTGGCG CGGCAAGCGG CATACTCGGT ACGCTGTTTT TCTGCTACGG
	351 CGCGGTGTGC ATGGCTTTGC CCGTATCCG TctGCAAAAC CGGCGCAACT
	401 ATGtgcCGT ATTTCGAATA TTTGTGCTGG GCGGTACGCA TGCGcgTTC
	451 CACGtccAgc tGCACACCG CAACCTAGGC GGACTCTTGA GCGGATTGCA
55	501 GTGGGCGCTG GTTATGCTGT CGGGCTTTAT CGGCTGATT GGGATGAGGA
	551 TTATTTCGTT TTTTACGTCC AAACGTTTGA ACGTCCGCGA GATTCCCAAT
	601 CCGAATGGG TGGGCGAGGC TCGCTGTGG CATACCATGC TGACCCCAT

-407-

5 651 ACTGATGGCG CACGGGCTGA TGCCTGGCT GTCGCGGCT TCGCGTTTG
 701 CGCGGGCGGT GATTTTTACC GTACAGGTGT ACCGCTGGTG GTATAAACCC
 751 GTATTGAAAG AACCGATGCT GTGGATTCTG TTTGCGCGCT ATCTGTTTAC
 801 CGGATTGGGG CTGATTGCGG TCGGCGGCTG TTATTTCAA CCGCTCTTCC
 851 TCAATCTGGG CGTACATCTG ATCGGGGCTG GCGGTATCGG CGTCTGACT
 901 TTGGGCGATGA TGGCGCGTAC CGGCTCGGT CATACGGCA ATTCGATTTA
 951 TCGCGCGCCC AAAGCCGCTT CCGTTCGGTT TTGGCTGATG ATGGCGCGAA
 1001 CGCGCGTCCG TATGGTTGCG GTATTTTCTT CCGGCACTGC CTACACGCAC
 1051 AGCATCCGCA CGTCTTCGGT TTTGTTTGCA CTCGCGCTGC TGGTGTATGC
 1101 GTGGAAATAC ATTCGCTGGC TGATCCGTCC CGGTCGGAC GGCAGGCCCG
 1151 GTTGA

This encodes a protein having amino acid sequence <SEQ ID 714; ORF47ng-1>:

1 MKFTKHPVMA MAFRPFYSLA ALYGALSVLL WGFYGTGTHE LSGFYWHAHE
 51 MIWGYAGLVV IAFLLTAVAT WTGQPTTGG VLVGLTAFWL AARIAEIPF
 101 WGAASGILG TLFFWVGAVC MALPVIRSQW RRYVAVFAI FVLGGTHAF
 151 HVQLHNGNLG GLLSGLQSL VMVSGFIGLI GMRIISFFTS KRLNVQPIPS
 201 PKWVAQASLW LPMLTAILKA HGVPWPLSAA FAFAGVIFT VOYVRWWYKP
 251 VLKEPMLWIL FAGYLTGLG LIAVGASYFK PAFNLGVHL IGVGIGSVLT
 301 LGMMARTALG HTGNSIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH
 351 SIRTSSVLEA LALLVYAWKY IPWLIRPRSD GRGP*

ORF47ng-1 and ORF47-1 show 97.4% identity in 384 aa overlap:

	orf47-1.pep	10	20	30	40	50	60
	MKFTKHPVMAAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFYWHAHEMIWGYAGLVV						
25	orf47ng-1	10	20	30	40	50	60
	MKFTKHPVMAAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFYWHAHEMIWGYAGLVV						
	orf47-1.pep	70	80	90	100	110	120
30	IAFLLTAVATWTGQPTTGGVLVGLTIFWLAARIAAFIPGWGASASGLTGLFFWYGAVC						
	orf47ng-1	70	80	90	100	110	120
	IAFLLTAVATWTGQPTTGGVLVGLTAFWLAARIAAFIPGWGASASGLTGLTGLFFWYGAVC						
	orf47-1.pep	130	140	150	160	170	180
35	MALPVIRSQWQRNYVAVFALEVLGGTHAFAFHVQLHNGNLGGLLSGLQSGLVMSVSGFIGLI						
	orf47ng-1	130	140	150	160	170	180
	MALPVIRSQWRNRNYVAVFAIFVLGGTHAFAFHVQLHNGNLGGLLSGLQSGLVMSVSGFIGLI						
	orf47-1.pep	190	200	210	220	230	240
40	GTIRISFFTSKRLNVQIPSPKWVAQASLWLPMLTAMLMAHGVLAWSAVFAFAAGVIFT						
	orf47ng-1	190	200	210	220	230	240
	GMRIISFFTSKRLNVQIPSPKWVAQASLWLPMLTAMLMAHGVMPWLSAAFAFAAGVIFT						
	orf47-1.pep	250	260	270	280	290	300
45	VOYVRWWYKPVLEKPEMLWILFAGYLTGLG LIAVGASYFKPAFNLGVHLIGVGIGSVLT						
	orf47ng-1	250	260	270	280	290	300
	VOYVRWWYKPVLEKPEMLWILFAGYLTGLG LIAVGASYFKPAFNLGVHLIGVGIGSVLT						
	orf47-1.pep	310	320	330	340	350	360
50	LGMMARTALGHTGNSIYPPKAVPVAFWLMAATAVRMVAVFSSGTAYTHSIRTSSVLEA						
	orf47ng-1	310	320	330	340	350	360
	LGMMARTALGHTGNSIYPPKAVPVAFWLMAATAVRMVAVFSSGTAYTHSIRTSSVLEA						
	orf47-1.pep	370	380				
60	LALLVYAWKYIPWLIRPRSDGRGPX						
	orf47ng-1	370	380				
	LALLVYAWKYIPWLIRPRSDGRGPX						

Furthermore, ORF47ng-1 shows significant homology to an ORF from *Pseudomonas stutzeri*:

gnl|PID|e246540 (273914) ORF396 protein [Pseudomonas stutzeri] Length = 396
 Score = 155 bits (389), Expect = 5e-37

-408-

Identities = 121/391 (30%), Positives = 169/391 (42%), Gaps = 21/391 (5%)

```

Query: 7  PWWAMAFRPFYSLAALYGALSLLWGFYGTGTHELSGFY-----WHAHEMHWYAGLV 59
P+W +AFRPF+ +LY L++ LW +TG GF WH HEM++G+A +
5  Sbjct: 14 PIWRIAFRPFPLAGSLYALLAIPWVAWATGLWP--GFQPTGGWLAWHREMLPGFAMAI 71

Query: 60  VIAFLLTAVATWTGQPPTGGVVLGLTAFWLAARIPAFIPWGAAASGILGLTFFWYGA 119
V FLLTAV TWTGQ G LVGL A WLAAR++ G AA L LF
10  Sbjct: 72  VAGFLLTAVQTWTGCTAPSGNRLVGLAAVWLAARL--GWLFGLPAWLAPLDLLFLVALVW 130

Query: 120  CMALPVIRSQNRNRYVAVFAIFVLGGTHAAFXXXXXXXXXXXXXXXXXXMSVSGIFL 179
MA + + +RNY V + ++ G +V+ + L
15  Sbjct: 131 MMAQMLWAVRQKRNYPVVVLSMLGADVLIITGLLQGNDAIQRQGVLAGLWLVLAALMAL 190

Query: 180  IGMRIISFPTSKRLNVPQIPSP-KWVAQASLWLPMLTALIMAHGV---MPWLSAAFAFA 234
IG R+I FET + L P W+ A L + +A+L A GV P L F A
15  Sbjct: 191 IGGRIVIFPTQRLGKVDVAVKFWWLDVALLVGTGVIALLHAFGVAMRQPQLGLLFV-A 249

Query: 235  AGVIPTQVYRWYKPVLEKPMWLIFLFGYLTGLGLIAGVASYF-KPAFXXXXXXXXXXXXX 293
GV +++ RW+ K + K +LW L L+ + + +F A
20  Sbjct: 250 IGVGHLLRMRWYDKGIWKVGLNLSLHVAMLLVVAAPGLALHFGLLAQSSPSLHALSV 309

Query: 294  XXXXXXXXXXXMARTALGHTGNSIYPPKAVPVAFWLXXXXXXXXXXXXXSSGTAYTHSIR 353
M+AR LGHTG + P + AF L F S +
25  Sbjct: 310 GMSGLILAMIAVRTLGHTGRPLQLPAGIIG-AFVL---FNLGTAARVFLSVANVPVGLW 365

Query: 354  TSSVLFALALLVYAWKYPWILIRPSDGRGP 384
++V + LA +Y +Y+ P L+ R DG PG
30  Sbjct: 366 LAACVTLAFALYVWRYPMLVAARVDGHPG 396

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 85

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 715>:

```

35      1  ..ATGCCGCTCTG AAGGTTTCAGA CGGCTCGGT GyCGGGGAy CAGAAGyGGT
51      AGCGCATGCC CAATGAGACT TCGTGGGTTT TGAAGCGGGT GTTTTCCAAG
101     CGTCCCGAGT TGTGGTAACG GTATCCGGTG TCyAaRGTA GCTTGGGyGT
151     GATGTCGAAa CCGACACCGG CGATGACACC AAGACCyAmG CTGCTGATrC
201     TGTGCTTTC GTGATAGGSA GdTTTyTGG kmksAsyTTG TArATwkkG
40      251  CCTsCwTG kAGmGCCCKtK KkyTGGTkKa swGrArTAG TCGTGGTtTy
301     TKTtTyCACc GAATGAACyT GATGTTTAAC GTGTCCGTAG GCGACGCGCG
351     CGCCGATATA GGGTTTGAAT TTATCGTTGA GTTTGAAATC GTAAATGGCG
401     GACAAAGCGGA GAGAAAGAAC GCGCTGGAAG CTGCGGTTTC CTGATGTTT
45      451  TGTtTGGGTT TCTTGTAGT TGTtGTTTAT CTCTTCAGTA ACTTTTTTGT
501     TAGAAGAATT ACTTCTCTTC CATTTTCTGT AACTGCGGTA ATCTGCGCGT
551     ATTCTCCAGC GCGCGAATTC ..

```

This corresponds to the amino acid sequence <SEQ ID 716; ORF67>:

```

50      1  ..MPSEGSdGXG XGEKXVAHA QXDFVGFEG VFGASPVVVT VSGVXXQLGX
51      DVETDTGGDT KTXAAXVAF VIGRFKXGXL YXKAXXXAX XWXXXSRGF
101     XXHMRMLMFN VSGVDARADI GFETVEFEI VNGGQARRN GVZAAVSLMF
151     CLGFFVVVVY LFSNFFSRRI TFFPFSVTGI ICrySPAEEI ..

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF67 shows 51.8% identity over 199 aa overlap with a predicted ORF (ORF67ng) from

55 *N.gonorrhoeae*:

	orf67.pep		MPSEGSDDGXGXGEXXVAHAQXDVFVGFAG	30
	orf67.ng	TNFIETIAVLVSGMTVRVFCARPAFVNGGRLEMPSEGSDDGIGIGISEVAHAQRGVGFVGFAG	146	
5		90 100 110 120 130 140		
	orf67.pep	VFQASPVVVTVSGVXXQLGXDVETDTGDDTKTXAADKXVAFVIGRFGXKLYXXAXXXXAX	90	
	orf67.ng	VFQASPVVVAVAGVQAGRDVYAHARHRAEAQAAAAVFLIGVFLRMSVRINRNCVSI	206	
10				
	orf67.pep	XWXXXSRGXGXXHRMNLMEVNSVGDARADIGFEFTVEFIVNGGQAERRNGVEAAVSLMF	150	
	orf67.ng	TRVGKSTCYFFSRIDAVSDVSVGDARTDIGFEFVVEFIVNGGQAERRNGVECAVLMF	266	
15				
	orf67.pep	CLGFFVV-----VVYLSNFSSRRITFE-PSVVTGILCRYSFAAEI	190	
	orf67.ng	RLLVFVTKLVAAKSFIIISQFLFYVHGIFIVVFPVVTGIRGDAPAAEVVADRHGPDGDM	326	

The ORF67ng nucleotide sequence <SEQ ID 717> is predicted to encode a protein comprising amino acid sequence <SEQ ID 718>:

20	1	MPSETVGSIV	NVGVDESVEG	SPFPFSIQHF	YRFHRIHRIR	IFRPPGPMQL
	51	NRHSHGSGNL	GRGWATVLS	DKFPCQVRI	PACAGMTNFE	IAVLSGMTVR
	101	VFYCARPAVP	NGGRLEMPSE	SGDGGIGES	EAVAHAQRGF	VGFEAGVFOA
	151	SPVVAVAGV	QCGAGRDVYA	HARHRAEAQA	AAVAFLIGV	FLRMSVRINR
	201	NCCVSITRVG	GKSTCYFFSR	IDAVSDVSVG	DARTDIGFEF	VVEFIVNGG
	251	QAERRNGVEC	AVFLMFRLLV	FYVKLVAASK	FIILSFQLEY	VHGIFIVVFP
25	301	PVTGIIRGDA	PAAEVVADRH	PGVDGMRTDV	SEIIAYRAYF	VFAWSGWFR
	351	IVGNAFGVG	*			

Based on the presence of a several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

30 Example 86

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 719>

	1	ATGTTTGCTT	TTTTAGAAGC	CTTTTTTGTC	GAATACGGTT	ATGCGGCTGT
	51	TTTTTTTGTA	TTGGTCATCT	CGGGTTTCGG	CGTGCCGATT	CCCGAGGATT
	101	TGACCTTGGT	AACAGCGCGC	GTGATTTCGG	GTATGGGTTA	TACCAATCCG
35	151	CATATTATGT	TTGCAGTCGG	TATGCTCGGC	GTATTGGTCC	GGGACGGCAT
	201	CATGTTCCGC	CGCGAGCGAA	TTTGGGGGCA	GAATTCCTA	CGTTCAACAC
	251	CTATTGCGG	CATCATGACG	CGGAACGTT	ATGAGCAGGT	TCAGGAAAAA
	301	TTGCAGCAAT	ACGGTAACTG	GGCTTATATT	GTGCGCCGTT	TCCTGCCCGG
	351	TTTGAGAAGC	CGCGTATTGT	TACAGCCGG	TATCAGCCGC	AAGGTTTCAT
40	401	ACTTGCGTTT	TATCATTATG	GATGGACTGG	CCGCA...	

This corresponds to the amino acid sequence <SEQ ID 720; ORF78>:

	1	MFAFLFAFFV	EYGYAARVFF	LVICGFGVPI	PEDTLIVTGG	VISGMGYTNP
	51	HIMFAVGMGL	VLVGDGIMFA	AGRTWGXLL	XFPIAXIMT	PKRYEQVQEK
	101	FDKYGNVVL	VARFLGLEAT	AVFVTAGISR	KVSYLRFIIM	DGLAA...

45 Further work revealed the complete nucleotide sequence <SEQ ID 721>:

	1	ATGTTTGCTT	TTTTAGAAGC	CTTTTTTGTC	GAATACGGTT	ATGCGGCTGT
	51	TTTTTTTGTA	TTGGTCATCT	CGGGTTTCGG	CGTGCCGATT	CCCGAGGATT
	101	TGACCTTGGT	AACAGCGCGC	GTGATTTCGG	GTATGGGTTA	TACCAATCCG
	151	CATATTATGT	TTGCAGTCGG	TATGCTCGGC	GTATTGGTCC	GGGACGGCAT
50	201	CATGTTCCGC	CGCGAGCGAA	TTTGGGGGCA	GAAATTCCTA	AGGTTCAAAAC
	251	CTATTGCGCG	CATCATGACG	CGGAACGTT	ATGAGCAGGT	TCAGGAAAAA
	301	TTGCAGCAAT	ACGGTAACTG	GGCTTATATT	GTGCGCCGTT	TCCTGCCCGG
	351	TTTGAGAAGC	CGCGTATTGT	TACAGCCGG	TATCAGCCGC	AAGGTTTCAT
	401	ACTTGCGTTT	TATCATTATG	GATGGACTGG	CCGCACTGAT	TTCCGTCCCT
55	451	ATTGGGATT	ATCTGGGCGA	ATACGCTGCG	CACAACTACG	ATTGGCTGAT

501 GCGGAAATG CACAGCTGC AATCGGGTAT TTTTGTATC TTGGGTATAG
 551 GTGCGACCGT TGTGCTTGG ATTTGTTGGA AAAACGCCA ACGTATCCAG
 601 TTTTACGCCA GCAATTTGAA AGAAACGGG GCGCAACGA AAGCCGCCAA
 651 GGCAGCCAAA AAGCCGCGC AAGCAACA ATAA

5 This corresponds to the amino acid sequence <SEQ ID 722; ORF78-1>:

1 MFAFLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMGYTNP
 51 HIMEFAVGMGLG VLVGDMIFFA AGRINGQKIL RFKPIARIMT PKRYEQVQEK
 101 FDKYGNWVLG VARFLPGLRT AVFVTAGISR KVSYLRFIIM DGLAALISVP
 151 IWIYLGEGYA HNIDWLMAKM HSLQSGIFVI LGIGATVVAW IWWKKRQRIQ
 201 FYRSKLKEKR AQRKAACA KAAQSKQ*

Computer analysis of this amino acid sequence predicts several transmembrane domains, and also gave the following results:

Homology with the dedA homologue of *H. influenzae* (accession number P45280)

ORF78 and the dedA homologue show 58% aa identity in 144aa overlap:

15 Orf78: 4 FLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGM--GYTNPHIMFAVGMGLG 61
 FL FF EYGY AV FVL+ICGFGVPIPED+TLV+GGVI+G+ N H+M V M+GV
 DedA: 20 FLIGFTTEYGYAVLFLVLIICGFGVPIPEDITLVSGGVIAGLYPENVNSHMLLVSMIGV 79
 20 Orf78: 62 LVGDMIFFAAGRIWGXQLXFXPIAXIMTPKRYEQVQEKFDKYGWNLVVARFLPGLRTA 121
 L GD M+ GRI+G L F PI I+T R V+EK+YGN VLFVARFLPGLR
 DedA: 80 LAGDSMYLWLGRIYGTILRFPIRIRIVTLQRLRMVREKFSQYGNRVLVVARFLPGLRAP 139
 Orf78: 122 VFVTAGISRKVSYLRFIIMDGLAA 145
 +++ +GI+R+VSY+RF+++D AA
 25 DedA: 140 IYMWSGITRRVSYRVFLIDFCAA 163

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF78 shows 93.8% identity over a 145aa overlap with an ORF (ORF78a) from strain A of *N. meningitidis*:

30 orf78.pep 10 20 30 40 50 60
 MFALFAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMGLG
 orf78a MFALFAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMGLG
 35 orf78.pep 70 80 90 100 110 120
 VLVGDMIFFAAGRIWGXQLXFXPIAXIMTPKRYEQVQEKFDKYGWNLVVARFLPGLRT
 orf78a VLVGDMIFFAAGRIWGXQLXFXPIAXIMTPKRYEQVQEKFDKYGWNLVVARFLPGLRT
 40 orf78.pep 130 140
 AVFVTAGISRKVSYLRFIIMDGLAA
 orf78a AVFVTAGISRKVSYLRFIIMDGLAALISVPWYIYLGEGAHNIDWLMAKMHSLSQSGIFIA
 45 orf78a 130 140 150 160 170 180

The complete length ORF78a nucleotide sequence <SEQ ID 723> is:

1 ATGTTTGCCC TTTTGAAGC CTTTTTGTG GAATACGGCT ATGCGGCCGT
 51 GTTTTTCGTT TTGTCATCT GCGGTTTCGG CGTGCCGATT CCGAGGATT
 101 TGACCTTTGGT AACAGCGCGC GTGATTTCGG GTATGGGTTA TACCAATCCG
 151 CATATTATGT TTGCACTCGG TATGCTCGGC GTATTGCTCG GGGACGGCAT
 201 CATGTTCCGCC GCCGACGCA TCTGGGGGCA GAAATCCTC AAGTTCACAA
 251 CGATTGGCGC CATCATGACG CCGAAACGTT ACGCACAGGT TCAGGAAAAA
 301 TTCGACAAAT ACGGCAACTG GGTGTATTAT GTGCTCGTTC TCCTGCCCGT
 351 TTTCGGGACT GCGGTTTCGG TTACGCGCGG CACACACCGC AAGTATCGT
 401 ATCTGCGCTT TCTGATTATG GACGGGCTTG CCGCGTGTAT TTCGTCGCGC
 451 GTTTGGATTT ACTTGGGCGA GTACGGCGCG CACACATCG ATTGCTGATG

```

501 GGCAGAAATG CACAGCCTGC AATCCGGCAT CTTCATCGCA TTGGGCGTGC
551 TGGCGGCGGC GCTGGCGTGG TTCTGGTGGC GCAACGCGC ACATTATCAG
601 CTTTACCGCG CACATTGTAG CGAAACGCG GCCAACGCA AGCGGAAAA
651 GGCAGCGAAA AAGCGGCAC AGAAGCAGCA GTAA

```

5 This encodes a protein having amino acid sequence <SEQ ID 724>:

```

1 MFALLEAFFV EYGYAAFFV LVICGFGVPI PEDLTFLTGG VISGMGYTNP
51 HMFVAVGMLG VLVGDGIMFA AGRINGQKIL KFKPIARIMT PKRYAQVQEK
101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRLIM DGLAALISVP
151 VWIYLGEYGA HNDWLMARMT HSLQSGIFIA LGVLAALAW FWRKRHRHYQ
10 LYRAQLSEKR AKRKAERKAA KAAQKQ*

```

ORF78a and ORF78-1 show 89.0% identity in 227 aa overlap:

```

10 20 30 40 50 60
orf78a.pep MFALLEAFFVEYGYAAFFVLVICGFGVPIPEDLTFLTGGVISGMGYTNPHIMFVAVGMLG
15 orf78-1 MFALLEAFFVEYGYAAFFVLVICGFGVPIPEDLTFLTGGVISGMGYTNPHIMFVAVGMLG
10 20 30 40 50 60
70 80 90 100 110 120
orf78a.pep VLVGDGIMFAAGRIWQKILKFKPIARIMTPKRYAQVQEKFDKYNWVLFVARFLPGLRT
20 orf78-1 VLVGDGIMFAAGRIWQKILRKFPIARIMTPKRYEQVQEKFDKYNWVLFVARFLPGLRT
70 80 90 100 110 120
130 140 150 160 170 180
orf78a.pep AVFVTAGISRKVSYLRLFLIMDGLAALISVPVWIYLGEYGAHNIDWLMARMTLSQSGIFIA
25 orf78-1 AVFVTAGISRKVSYLRLFLIMDGLAALISVPVWIYLGEYGAHNIDWLMARMTLSQSGIFV
130 140 150 160 170 180
190 200 210 220
orf78a.pep LGVLAALAWFWRKRHRHYQLYRAQLSEKRAKRAKAAKAAQKQX
30 orf78-1 LGIGATVVAWIKKRRQIQIFRSKLKEKRAKRAKAAKAAQKQX
190 200 210 220
35

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF78 shows 97.4% identity over 38 aa overlap with a predicted ORF (ORF78ng) from *N. gonorrhoeae*:

```

40 orf78.pep XXLFXPIAXIMTPKRYEQVQEKFDKYNWVLFVARFLPGLRTAVFVTAGISRKVSYLRF 137
orf78ng YPVLVARFLPGLRTAVFVTAGISRKVSYLRF 32
orf78.pep IIMDGLAA 145
45 orf78ng IIMDGLAALISVPVWIYLGEYGAHNIDWLMARMTLSQSGIFIALGVLAALAWFWRKR 92

```

The ORF78ng nucleotide sequence <SEQ ID 725> is predicted to encode a protein comprising amino acid sequence <SEQ ID 726>:

```

1 .YPVLVARFL PGLRTAVFT AGISRKVSYL RFLIMDGLAA LISVPVWIYL
51 GEYGAHNIDW LAKRMHLSQS GIFIALGVLA AALAWFWRK RRYHYQLYRAQ
50 101 LSEKRAKRA EKAAKKAAQK Q*

```

Further work revealed the complete gonococcal nucleotide sequence <SEQ ID 727>:

```

1 atgtttgccc tttTggaagc CTTTTTGTGC GAATacggCt atgcGGCGGT
51 GTTTTTCGTT TTGTCATCT GCGGTTTCGG CGTGCCGATT CCGCAAGATT
101 TGACCTTTGT AAGCGGCGCG GTGATTTCGG GTATGGGTTA TACCAATCCG
55 151 CATATTATGT TTGCGGTCGG TATGCTCGCG GTGTTCGGCG CGCAGCGCGT
201 GATGTTTGCC GCGGAGCGCA TCTGGGGGCA GAAATCTTC AAGTTCAAAC
251 CGATTGCGCG CATCATGACG CCGAAGCTT ACGCGCAGT TCAGGAAAAA
301 TTCGACAAAT ACGGCAACTG GGTTCTGTTT GTCGCCCGTT TCCTGCCGGG

```

5

351 TTTGGGACAT GCCGTTTTCG TTACGCGCGG CATCAGCCGC AAGATATCGT
401 ATCTCGGCTT TCTGATTATG GAGCGGCTCG CCGCGCTGAT TTCCGTCGCC
501 TTGTGGATTG ATCTGGGCGA GATCGGCGCG CACACATCAT TTCTGCTGAT
601 GGCGAAMATT CACAGCTTCG AATCGGCATC CTTCATCGCA TTGGGCGTGG
701 TGGCGCGCGC CTGCGGCTCG TCTCGTGTCG GCRAACACGC ACATTATCAG
801 CTTCACGCGC CACAATTGAG GCAAAACGCG GTCAACACGA AGGCGGAAAA
901 GGCAAGCGAA AGAGCGGCAC AGAGACGACA GTTaa

This corresponds to the amino acid sequence <SEQ ID 728; ORF78ng-1>:

10

1 MFALLEAFFV EYGAAVFFV LVICGFGVPI PEDLTLTGG VISGMYTNP
51 HIMFAVGMIG VLADGVMFA AGRIWGQKIL KFKPIARIMT PKRYAQVQEK
101 FDKYGNVYLF VARFLPGLR AVEFTAGISR KVSYLRLIM DGLAALISVP
151 VWIIYLGEYA HNIDWIMAKM HSLOSGTISA LGVLAAALR FWWRKRRHYQ
201 LYRAOLSEKR AKRAEAKAK KAAOKCO*

ORF78ng-1 and ORF78-1 show 88.1% identity in 227 aa overlap:

15

orf78-1.pep MF¹⁰AFLEAFFVEYGYAAVFFVLVICGFGVP|PEDLT|LVTTGGVISGMGYTNPHIMFAVGMLG
|||:::
orf78nq-1 MF¹⁰AFLEAFFVEYGYAAVFFVLVICGFGVP|PEDLT|LVTTGGVISGMGYTNPHIMFAVGMLG

20

orf78-1. pep VLVGDIIMFAAGRIWGQKILRFKPIARIMTPKRYEQVQEKDFKYGNWLVFVARFLPGLRT

25

orf78ng-1 VLAGDGVMEFAAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLEFVARFLPGLRT

30

```

      130      140      150      160      170      180
orf78-1.pep  AVFVTAGISRKVSYLRFITMDGLAALISVPITWYLYGEGAHNIDWLMAKMHSLSQSGIFVI
orf78ng-1    AVFVTAGISRKVSYLRFITMDGLAALISVPVWYLYGEGAHNIDWLMAKMHSLSQSGIFTA
      130      140      150      160      170      180

```

35

```

190      200      210      220
orf78-1.pep LGIGATVVAIWWKKRQRTQFYRSKLKEKRAAKAAKAAQSKQX
||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
orf78ng-1    LGVLAALAAWFWWRKKRRHYQLRYALQSEKRAKRAKAAKAAQKQX
190      200      210      220

```

Furthermore, orf78ng-1 shows homology to the dedA protein from *H. influenzae*:

40

sp|P45280|YG29_HAEIN HYPOTHETICAL PROTEIN HI1629 >gi|1073983|pir|ID64133 dedA protein (dedA) homolog - Haemophilus influenzae (strain Rd KW20)
>gi|1574476 (U32936) dedA protein (dedA) [Haemophilus influenzae] Length = 212
Score = 223 bits (563), Expect = 7e-58
Identities = 108/182 (59%), Positives = 140/182 (76%), Gaps = 2/182 (1%)

45

Query: 5 LEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLTVTGGVISGM--GYTNPHIMFAVGMGLVL 62
L FF EYGY AV FVL+ICGFGVPIPED+TLV+GGVI+G+ N H+M V M+GVL
Shift: 21 LIGFTEYGYAAVLEVLIIICGFGVPIPEDITLVSGGVIAGLYPENNVSHIMLLVSMIGVL 80

50

Query: 63 AGDGVMF AAGRIWGQKILFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRTAV 122
AGD M+ GRI+G KIL+F+PI RI+T +R V+EKF +YGN VLFVARFLPGLR +
Sbjct: 81 AGDSCHYWLGR IYGTKILFRPIRRIVTLQRLRMVREKFSQYGNRVLFVARFLPGLRAPI 140

55

Query: 123 FVTAGISRKVSYLRFLIMDGLAALISVPPWIYLGEGAHNIDWLMAKMHSLSQSGIFIALG 182
 ++ +GI+R+VSY+RF+++D AA+ISVP+WIIYLG EG A N+DWL ++ Q I+I +G
 Sbjct: 141 YMVSGITRRVSYVRFVLIDFCAAIISVPIWIYLGELGAKNLDWLHTQIQKGQIVIIYIFIG 200

Query: 183 VL 184
L
Sbjct: 201 YL 202

60

Based on this analysis, including the presence of putative transmembrane domains, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 87

- 5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 729>:

```

1 ATGAAAAAAT TATTGGCGGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
51 TTCCGCCGCC GGAGTCCACG TTGAGGACGG CTGGCGCGCG ACCACGCTCG
101 AAGGTATGAA AATAGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC
151 AAACAAGACT TTTTGTCCGG CGGAAGCAGC CCGGTTGCGC ACCGCTCGA
201 AGTGCAATAC CACATCAACG ACAACGCGGT GATGCGGATG CGCGAAGTGC
251 AAGGCGCGGT GCCTTTGGAA GCGAAATCCG TTACCGAAGT CAAACCCGCG
301 AGCTATCATG TGATGTTTAT GGGTTTGAAG AAACAATTAA AAGAGGGCGA
351 TAAATTCGCC GTTACCCCTGA AATTTAAAAA CGCCAAAGCG CAAACCGTCC
401 AACTGGAAGT CAAATCGCG CCGATGCCGG CAATGAACCA C...
```

- 15 This corresponds to the amino acid sequence <SEQ ID 730; ORF79>:

```

1 MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKIGG AFMKIHNDDEA
51 KQDFLLGGSS PVADRVVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
101 SYHVMFMGLK KQLKEGDKIP VTLKFNKA QTVQLEVKIA PMPAMNH...
```

Further work revealed the complete nucleotide sequence <SEQ ID 731>:

- 20
- ```

1 ATGAAAAAAT TATTGGCGGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
51 TTCCGCCGCC GGAGTCCACG TTGAGGACGG CTGGCGCGCG ACCACGCTCG
101 AAGGTATGAA AATAGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC
151 AAACAAGACT TTTTGTCCGG CGGAAGCAGC CCGGTTGCGC ACCGCTCGA
201 AGTGCAATAC CACATCAACG ACAACGCGGT GATGCGGATG CGCGAAGTGC
251 AAGGCGCGGT GCCTTTGGAA GCGAAATCCG TTACCGAAGT CAAACCCGCG
301 AGCTATCATG TGATGTTTAT GGGTTTGAAG AAACAATTAA AAGAGGGCGA
351 TAAATTCGCC GTTACCCCTGA AATTTAAAAA CGCCAAAGCG CAAACCGTCC
401 AACTGGAAGT CAAATCGCG CCGATGCCGG CAATGAACCA CGGTCATCAC
451 CACGCGCAAG CGCATCAGCA CTAA
```

- 30 This corresponds to the amino acid sequence <SEQ ID 732; ORF79-1>:

```

1 MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKIGG AFMKIHNDDEA
51 KQDFLLGGSS PVADRVVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
101 SYHVMFMGLK KQLKEGDKIP VTLKFNKA QTVQLEVKIA PMPAMNHGHH
151 HGEAHQH*
```

- 35 Computer analysis of this amino acid sequence revealed a putative leader peptide and also gave the following results:

### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF79 shows 94.6% identity over a 147aa overlap with an ORF (ORF79a) from strain A of *N.meningitidis*:

- 40
- ```

orf79.pep MKKLLAAVMMAGLAGAVSAAGVHVEDGWAR TTVEGMKIGGAFMKIHNDDEAKQDFLLGGSS
orf79a MKKLLAAVMMAGLAGAVSAAGIHVEDGWAR TTVEGMKIGGAFMKIHNDDEAKQDFLLGGSS
10 20 30 40 50 60
```
- 45
- ```

orf79.pep PVADRVVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
orf79a PVADRVVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
70 80 90 100 110 120
```
- 50
- ```

orf79.pep
orf79a
70 80 90 100 110 120
```

```

      130      140
orf79.pep VTLKFKNAKAQTVQLEVKIAPMPAMNH
5         orf79a VTLKFKNAKAQTVQLEVKIAPMSAMDHGHHGEAHQH
      130      140      150

```

The complete length ORF79a nucleotide sequence <SEQ ID 733> is:

```

      1 ATGAANAAC TATTGGCAGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
51 TTCCGCGGCC GGAATCCAGC TTGAGGACGG CTGGGCGCGC ACCACCGTCG
101 AAGGTATGAA AATGGCGCGC GCGTTCATGA AAATCCACAA CGACGAAGCC
151 AAACAAGACT TTTTGTCTCG CGGAAGCAGC CCTGTTCGCG ACCCGGTCGA
201 AGTCGATACC CATATCAATG ATACCGGTGT GATGCGGATG CGCGAAGTCG
251 AAGGCGCGCT GCCTTTGGAG GCGAATCCG TTACCGGACT CAAACCCGCG
301 AGCTATCATG TCATGTTTAT GGGTNTGAAA AAACAATTAA AAGANGCGGA
15 351 CAAGATTCCC GTTACCTCGA AATTAAAAA CGCCAAAGCA CAACCGCTCC
401 AACTGGAGGT CAACAACCGC CGGATGTCGG CAATGGACCA CGGTATCATC
451 CACGGCGAAG CGCATCAGCA CTA

```

This encodes a protein having amino acid sequence <SEQ ID 734>:

```

      1 MKYLLAAVMM AGLAGAVSAA GIHVEDGWAR TTVEGKMGG AFMKIHND
10 51 KQDFLLGSS PVADRVEVHT HINDNGVMRM REVEGGVPLE AKSVTELPK
101 SYHVMFMGKK KQLKXGDKIP VTLKFKNAKA QTVQLEVKTA PMSAMDHGHH
151 HGEAHQH*

```

ORF79a and ORF79-1 show 94.9% identity in 157 aa overlap:

```

      10      20      30      40      50      60
25 orf79a.pep MKKLLAAVMMAGLAGAVSAGIHHVEDGWARTTVEGKMGGAFMKIHNDKQDFLLGSS
      11
orf79-1 MKKLLAAVMMAGLAGAVSAGVHHVEDGWARTTVEGKMGGAFMKIHNDKQDFLLGSS
      10      20      30      40      50      60

      70      80      90      100      110      120
30 orf79a.pep PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELPKGSYHVMFMGKKQLKXGDKIP
      11
orf79-1 PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELPKGSYHVMFMGKKQLKEGDKIP
      70      80      90      100      110      120

      130      140      150
35 orf79a.pep VTLKFKNAKAQTVQLEVKIAPMSAMDHGHHGEAHQH
      11
orf79-1 VTLKFKNAKAQTVQLEVKIAPMPAMNHGHHGEAHQH
      130      140      150
40

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF79 shows 96.1% identity over 76 aa overlap with a predicted ORF (ORF79ng) from *N.gonorrhoeae*:

```

45 orf79.pep FMKIHNDKQDFLLGSSPVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELPKGS 101
      11
orf79ng INDNVMMRMREVEGGVPLEAKSVTELPKGS 30

      130      140      150
50 orf79.pep YHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEVKIAPMPAMNH 147
      11
orf79ng YHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEVKIAPMSAMNHGHHGEAHQH 86

```

An ORF79ng nucleotide sequence <SEQ ID 735> was predicted to encode a protein comprising amino acid sequence <SEQ ID 736>:

```

      1 ..INDNGVMRM EVKGGVPLEA KSVTELPKGS YHVMFMGLKK QLKEGDKIPV
55 51 TLKFKNAKAQ TVQLEVKIAP MSAMNHGHH GEAHQH*

```

Further work revealed the complete gonococcal DNA sequence <SEQ ID 737>:

```

1 ATGAAAAAAT TATTGGCAGC CGTGTATGAT GCAGGTTTGG CAGGCGCGGT
51 TTccgcgcgc GgagTccATG TCGAggACGG CTGGGCGGCG accaCTgtgc
101 aaggtATgaa aatggCGCGG GCytlCATat aaATCCACAA CGACGaaGcc
151 atacaaGACT ttgtgcTCgg CGGagcatg cccgttgcgc accgcGTCGA
201 AGTGCAtaca cacATCAACG ACAACGCGGT GATGCGGTAT CGCGAAGTCA
251 AAGGCGCGGT GCCTTTGGAG GCGAAATCCG TTACCGAACT CAAACCGGCG
301 AGCTATCACT TGATGTTTAT GGGTTTGAaa AAACAACtGA AAGAGGGCGCA
351 CAAGATTCCC GTTACCCTGA AATTTAAAAA CGCCAAAGCG CAAACGCTCC
401 AACTGGAAGT CAAAACCGCG CCGATGTCGG CRATGAACCA CGGTCATCAC
10 451 CACGCGGAAG CGCATCAGCA CTAa

```

This corresponds to the amino acid sequence <SEQ ID 738; ORF79ng-1>:

```

1 MKKLLAAVMM AGLAGAVSRA GVHVEDGWAR TTVEGMMKMG AFMKIHND EA
51 IQDFVLGGSM PVADRVVHT HINDNGVMRM REVKGGVPLE AKSVTELKPG
101 SYHVMFMGLK KQLKEGDKIP VTLKFNKAQA QTVQLEVKTA FMSAMNHGHH
15 151 HGEARQH*

```

ORF79ng-1 and ORF79-1 show 95.5% identity in 157 aa overlap:

```

                10      20      30      40      50      60
orf79-1.pep  MKKLLAAVMMAGLAGAVSAAAGVHVEDGWARTTVEGMKIGGAFMKIHND EAKQDFVLGGSS
                |||
20 orf79ng-1  MKKLLAAVMMAGLAGAVSAAAGVHVEDGWARTTVEGMKMGGA FMKIHND EAIQDFVLGGSM
                10      20      30      40      50      60

                70      80      90      100     110     120
orf79-1.pep  PVADRVVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
                |||
25 orf79ng-1  PVADRVVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
                70      80      90      100     110     120

                130     140     150
orf79-1.pep  VTLKFNKAQAQTVQLEVKTA FMSAMNHGHHHGEARQH
                |||
orf79ng-1    VTLKFNKAQAQTVQLEVKTA FMSAMNHGHHHGEARQH
                130     140     150

```

Furthermore, ORF79ng-1 shows significant homology to a protein from *Aquifex aeolicus*:

```

35 gi|2983695 (AE000731) putative protein [Aquifex aeolicus] Length = 151
   Score = 63.6 bits (152), Expect = 6e-10
   Identities = 38/114 (33%), Positives = 58/114 (50%), Gaps = 1/114 (0%)

40 Query: 24 VEDGWARTTVEGMKMGGA FMKIHND EAIQDFVLGGSM PVADRVVHTHINDNGVMRMREV 83
   V+ W G M I N+ D+++G +A RVE+H + +N V +M
   Sbjct: 27 VKHPVWMEPPPGPNTMMGMIIIVNEGDEPDYLI GAKTDIAQRVELHKTVIENDVAKMVPQ 86

   Query: 84 KGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIPVTLKFNKAQAQTVQLEV 137
   + + + K E K YHVM +GLKK++KEGDK+ V L F+ + TV+ V
45 Sbjct: 87 ER-IEIPKGVFEFKHGHGYHVMII GLKKRIKEGDKVKEVLI FEKSGKITVEAPV 139

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF79-1 (15.6kDa) was cloned in the pET vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 18A shows the results of affinity purification of the His-fusion protein. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 18B) These experiments confirm that ORF79-1 is a surface-exposed protein, and that it is a useful immunogen.

		70	80	90	100	110	120
orf98.pep		GFNIPGLGVIAIAVLFTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSEYVL					
5	orf98a	GFNIPGLGVIAIAVLFTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVKSLL					
		70	80	90	100	110	120
		130	140	150	160	170	180
10	orf98.pep	SDSSRSFKTPVLVFPFPGIWTIAFVSGQVSNVAKAALPKDGDYLSVYVPTTNPPTGGYY					
	orf98a	SDSSRSFKTPVLVFPFPGIWTIAFVSGQVSNVAKAALPKDGDYLSVYVPTTNPPTGGYY					
		130	140	150	160	170	180
		190	200	210	220	230	
15	orf98.pep	IMVKKSDVRELDMSVDLKYVISLGMVIPDDLVPKTLAKMPSEKADLPEQQX					
	orf98a	IMVKKSDVRELDMSVDLKYVISLGMVIPDDLVPKTLAKMPSEKADLPEQQX					
		190	200	210	220	230	

The complete length ORF98a nucleotide sequence <SEQ ID 743> is:

20	1	ATGACGGAAC	CTGCGGCCGA	AGGCGGCAAA	GCTGCGCAAG	CGTTAAAAAA
	51	ATATCTGATT	ACGSGCATT	TGGTCTGGCT	GCGGATGCG	GTAACGGTTT
	101	GGGTGGTTTC	CTATATCGTT	TCCGCGTCCG	ATCAGCTCGT	CAACTCGTGT
	151	CCGAAGCAAT	GGCGGCCGA	ATATGTTTGT	GGGTTTAATA	TCCGCGGCT
25	201	GGGCGTTATC	GTTGCCATTG	CCGTATTGTT	TGTAACCGGA	TTATTTCGCG
	251	CAAACGTATT	GGGCGGCCAG	ATTCTTTCGCG	CGTGGGACAG	CTTGTGGGG
	301	CGGATTCCGG	TGTGAAGTC	CATCTATTCTG	AGTGTGAAAA	AAGTATCCGA
	351	NTCGTTGCTG	TCCGACAGCA	GCGGTTGCTT	TAAACACCA	GTACTCGTGC
	401	CGTTTCCCCA	ATCGGGTATT	TGGACAATCG	CATTCTGTCT	CGGTCAAGTG
	451	TGCAATGCGG	TAAAGCCCG	ATTGCGGAAG	GACGCGGATT	ATCTTTCGCT
30	501	GTATGTTCCG	ACCAACCCGA	ATCCGACCGG	CGGTACTATT	ATATGTGTA
	551	AGAAAAGCGA	TCTGCCGAA	CTCGATTGA	CGGTGACCA	AGCGTGAAA
	601	TATCTGATT	CGTGGGTAT	GCTCATCCCT	GACGACTCGC	CGCTCAAAAC
	651	ATTGGCAGGA	CCTATGCCGT	CTGAAAAGGC	GGATTGCCCC	GAACAACAAT
	701	AA				

35 This encodes a protein having amino acid sequence <SEQ ID 744>:

	1	NTEPAABEGGK	AAKALKKYLI	TGILVNLPIA	VTWVVSYIV	SASDQVLNLL
	51	PKQNRPOYVL	GFNIPGLGVI	VIAIAVLFTG	LFAANVLGRQ	ILAAWDSLLG
	101	RIPVVKSIYS	SVKKVXSLL	SDSSRSFKTP	VLVFPQSGI	WTIAFVSGQV
40	151	SNVAKAALPK	DGDYLSVYVP	TTNPPTGGYY	IMVKKSDVRE	LMSVDLAK
	201	YVISLGMVIP	DDLVPKTLAG	PMPSEKADLE	EQQ*	

ORF98a and ORF98-1 show 98.7% identity in 233 aa overlap:

		10	20	30	40	50	60
orf98a.pep		NTEPAABEGGKAAKALKKYLTGILVNLPIAVTWWVVSYIVSASDQVLNLLPKQWRPOYVL					
45	orf98-1	NTEXAABEGGKAAKALKKYLTGILVNLPIAVTWWVVSYIVSASDQVLNLLPKQWRPOYVL					
		10	20	30	40	50	60
		70	80	90	100	110	120
50	orf98a.pep	GFNIPGLGVIAIAVLFTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVKSLL					
	orf98-1	GFNIPGLGVIAIAVLFTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVKSLL					
		70	80	90	100	110	120
		130	140	150	160	170	180
55	orf98a.pep	SDSSRSFKTPVLVFPFPGIWTIAFVSGQVSNVAKAALPKDGDYLSVYVPTTNPPTGGYY					
	orf98-1	SDSSRSFKTPVLVFPFPGIWTIAFVSGQVSNVAKAALPKDGDYLSVYVPTTNPPTGGYY					
		130	140	150	160	170	180
		190	200	210	220	230	
60	orf98a.pep	IMVKKSDVRELDMSVDLKYVISLGMVIPDDLVPKTLAGMPSEKADLPEQQX					
	orf98-1	IMVKKSDVRELDMSVDLKYVISLGMVIPDDLVPKTLAGMPSEKADLPEQQX					
		190	200	210	220	230	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF98 shows 95.3% identity over a 233 aa overlap with a predicted ORF (ORF98ng) from *N.gonorrhoeae*:

		10	20	30	40	50	60	
5	orf98.pep	MTVTAAGGKAALKKYLITGILVWLP	IAVTWVVSYIVSASDQVLNLLPKQWRPQYVL					60
	orf98ng	MTEPPAEGGKAALKKYLITGILVWLP	IAVTWVVSYIVSASDQVLNLLPKQWRPQYVL					60
10	orf98.pep	GFNIPGLGVIAIAVLFTVGLFAANVLGRQILAANDSL	LLGRIPVVKSIYSSVKKVSEYVL					120
	orf98ng	GFNIPGLGVIAIAVLFTVGLFAANVLGRQILAANDSL	LLGRIPVVKSIYSSVKKVSESL					120
	orf98.pep	SDSSRSFKTFVLVFPQSGIWTIAFVSGQVSNVKAAL	PXGDGDLVSVYVPTTNPPTGGY					180
	orf98ng	SDSSRSFKTFVLVFPQSGIWTIAFVSGQVSNVKAAL	PXGDGDLVSVYVPTTNPPTGGY					180
15	orf98.pep	IMVKKSDVRELDMSVDKLYVILSGMVI	PDDLVPKTLAXPMPSEKADLEPQQ					233
	orf98ng	IMVKKSDVRELDMSVDKLYVILSGMVI	PDDLVPKTLAXPMPPEKADLEPQQ					233

- 20 The complete length ORF98ng nucleotide sequence <SEQ ID 745> is predicted to encode a protein having amino acid sequence <SEQ ID 746>:

	1	MTEPPAEGGK	AAKALKKYL	ITGILVWLP	IAVTWVVSYIV	SASDQVLNLL
	51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLFTV	GLFAANVLGRQ	ILAANDSL
	101	RIPVVKSIYS	SVKKVSESL	SDSSRSFKTF	VLVFPQSGI	WTIAFVSGQV
25	151	SNVAKAALPQ	DGDLVSVYVP	TTNPPTGGY	IMVKKSDVRE	LDMSVDEALK
	201	YVILSGMVI	PDDLVPKTL	AXPMPPEKALP	EQQ*	

Further work revealed the complete nucleotide sequence <SEQ ID 747>:

	1	ATGACGGAAAC	CTGCGGCCGA	AGGCGGCAAA	GCTGCCAAGG	CGTAAAAAAA
	51	ATATCTGATT	ACAGGCAATT	TGGTCGGCT	CGCATTTGC	GTAACGGTTT
	101	GGGTGCTTC	CTATATCGTT	TCGCGCTCG	ACCAGCTTGT	CAACCTCGTG
30	151	CGGAAGCAAT	GCGCGGCCGA	ATATGTTTGT	GGGTTAATA	TCCCGGGCT
	201	CGGCGTTATT	GTTCACATG	CGGTATTGTT	TGTAACCGCA	TTATTGCGC
	251	CAAACTGTT	GGGCGGCCAG	ATTCTTGC	CGTGGACAG	CCTGTgggg
	301	cggaTTCCG	TTGTCAATC	CATCTATTG	AGTGTGAAA	AAGTATCCGA
35	351	ATCGCTGCTG	TCCGACAGCA	GCGGTTGTT	TAAAACGCG	GTACTCGTGC
	401	CGTTTCCCA	ATCGGGTATT	TGGACAATCG	CATTCTGTG	CGGTGAGTG
	451	TGCAATGCG	TAAAGCGCG	ATTGCGCGAG	GATGCGGATT	ATCTTCCGT
	501	GTATGTCGCG	ACCACGCCCA	ACCCGACCG	CGGTACTAT	ATTATGGTAA
	551	AGAAAAGCA	TGTGCGCGAA	CTCGATATGA	CGGTGGACGA	AGCGTTGAAA
40	601	TATGTGATT	CGCTGGGTAT	GTCATCCCT	GACGACCTCG	CCGTCAAAAC
	651	ATTGGCAGCA	CCTATGCCGC	CTGAAAAGCG	GGAGTTGCC	GAACACAAAT
	701	AA				

This corresponds to the amino acid sequence <SEQ ID 748; ORF98ng-1>:

	1	MTEPPAEGGK	AAKALKKYL	ITGILVWLP	IAVTWVVSYIV	SASDQVLNLL
45	51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLFTV	GLFAANVLGRQ	ILAANDSL
	101	RIPVVKSIYS	SVKKVSESL	SDSSRSFKTF	VLVFPQSGI	WTIAFVSGQV
	151	SNVAKAALPQ	DGDLVSVYVP	TTNPPTGGY	IMVKKSDVRE	LDMSVDEALK
	201	YVILSGMVI	PDDLVPKTL	AXPMPPEKALP	EQQ*	

ORF98ng-1 and ORF98-1 show 97.9% identity in 233 aa overlap:

		10	20	30	40	50	60	
50	orf98-1.pep	MTEPPAEGGKAALKKYLITGILVWLP	IAVTWVVSYIVSASDQVLNLLPKQWRPQYVL					
	orf98ng-1	MTEPPAEGGKAALKKYLITGILVWLP	IAVTWVVSYIVSASDQVLNLLPKQWRPQYVL					
55		70	80	90	100	110	120	
	orf98-1.pep	GFNIPGLGVIAIAVLFTVGLFAANVLGRQILAANDSL	LLGRIPVVKSIYSSVKKVSESL					

-419-

5	orf98ng-1	GFNIPGLGVIAIAVLFTVGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSESL	70	80	90	100	110	120
	orf98-1.pep	SDSSRSFKTPVLVPFPQGIWNTIAFVSGQVSNVAKALPKDGDYLSVYVPTTNPNTGGYY	130	140	150	160	170	180
10	orf98ng-1	SDSSRSFKTPVLVPFPQGIWNTIAFVSGQVSNVAKALPKDGDYLSVYVPTTNPNTGGYY	130	140	150	160	170	180
	orf98-1.pep	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPKTLAGMPSEKADLPEQQX	190	200	210	220	230	
15	orf98ng-1	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPKTLAGMPSEKADLPEQQX	190	200	210	220	230	

Based on this analysis, including the fact that the putative transmembrane domains in the gonococcal protein are identical to the sequences in the meningococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 89

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 749>:

20	1	ATGAAAACGG	TAGTCTGGAT	TGTCGTCCTG	TTTGGCCGCG	CCGTCGGACT
	51	GGCGCTGGCT	TCGGGCATT	ACACCGCGCA	CGTGTATATC	GTACTCGGAC
30	101	AGACCATGCT	CAGAATCAAC	CTGCACGCT	TTGTGTTAGG	TTGCTGATT
	151	GGCGCTGGT	TGTGGTATTT	CTTGTTTAA	TTCAATATCG	GggTACTCA
40	201	ATATCCCCGA	AAAGATGACG	CGTTTCGGTT	CGGcncGTAA	AGGCCkCAAG
	251	ssCGsGCTTG	CCTTGACAA	GGCGGGTTTG	GCATTTT	AAGCGGCTTT
50	301	TGAAAAGGCG	GACTAGAAAG	CCTCAGCGT	GTTGCTCAAC	AAAGTAGGCC
	351	GgAGAGCAAC	CGGACTTTTG	CATTGATGCT	GrGCGCGCAC	CGCGCCGAGC
60	401	AGATGGAAAA	CATCGAasTG	CGCGACCGTT	ATCTTCGCGA	AATCGCCAAA
	451	CTGCGCGAAA	AACAGCAGCT	TTCCGCTTAT	CTTTTGTG	CGGAATCGCG
70	501	GTTGAACCGG	CGCGATTACG	AAGCGCGGGA	AGCCAATCTT	CATGCGCGCG
	551	CGAAGATGAA	TGCCAACCTT	ACGCGCCTCG	TGCGTCTGCA	.ATTTCGTTAC
80	601	GCTTTCGACA	GGGCGCAGCG	GTTGCAAGTT	CTGGCAAAA	CGRAAAACT
	651	TTCCAAAGCG	GGCGCTTGG	CGAAATCGA	AATGSAACGG	TATCAAAATT
90	701	GGGCAATATG	GTGCGACGCT	GGCGATGCT	GCCGATCGCG	CGCTTTGAA
	751	AACCTGCTTG	AACGCGATT	CCGACAGCT	CAAAAACGGG	GAATTGAGCG
100	801	TATCGGTTGC	GGAAGAATAC	GAACGTTTGG	GACTGTATCG	CGATGCGGTC
	851	AAATGGGTCA	AACAGCATTA	TCCGCAAAAC	CGCGCCCGCG	AGCTTTTGGG
110	901	AGCCTTTGTC	GAAAGCGTGC	GCTTTTGGG	CGAGCGCGAA	CAGCAGAAAG
	951	CCATCGATT	TGCCGATGCT	TGGCTGAAAG	AACAGCCCGA	TACCGCGCTT
120	1001	CTGCTGATGT	ATCTCGGTGC	GCTCGCCTTC	GGCGCGAAAC	TTTGGGGCAA
	1051	GGCAAAAGGC	TACCTTGAAG	CGAGCATTGC	ATTAAAGCCG	AGTATTTCGG
130	1101	CGCGTTTGTG	TCTAACAAAG	GTTTTCGAGC	AAATCGGAGA	ACCGCAGAAG
	1151	CGCGAGGCGC	AC...			

This corresponds to the amino acid sequence <SEQ ID 750; ORF100>:

50	1	MKTVVVIVVL	FAAVGLALA	SGIYTGVDVY	VLGQTMRLIN	LHAFVLGSLI
	51	AVVVWYFLFK	FIIVGLNIPE	KMQRFGSARK	GXIOXLALNK	AGLAYFEGRF
60	101	EKAELEASRV	LVNKGVRDNR	TLALMLXAHA	AGQMNIXXR	DRLYAEIARL
	151	PEKQQLSRYL	LLAESALNRR	DYEAENANLH	AAAKMNNANLT	RLRLXIRYA
70	201	FDRGDALQVL	AKTEKLSKAG	ALGKSEMERY	QNWYRRQLA	DAADAAALKT
	251	CLKRIPDSLK	NGELSVSVAE	KYERLGLYAD	AVKWKQHYLP	XNRRFELLEA
80	301	FVESVRFELGE	REQQKAIDFA	DANLKEQFDN	ALLMLYGLRL	AFGRKLWGRA
	351	KGYLEASIAL	KFISIALRLVL	TKVFEIDGEF	QKAEH...	

Further work revealed the complete nucleotide sequence <SEQ ID 751>:

55	1	ATGAAAACGG	TAGTCTGGAT	TGTCGTCCTG	TTTGGCCGCG	CCGTCGGACT
	51	GGCGCTGGCT	TCGGGCATT	ACACCGCGCA	CGTGTATATC	GTACTCGGAC
60	101	AGACCATGCT	CAGAATCAAC	CTGCACGCT	TTGTGTTAGG	TTGCTGATT
	151	GGCGCTGGT	TGTGGTATTT	CTTGTTTAA	TTCAATATCG	GgTACTCA

201 TATCCCCGAA AAGATGCAGC GTTTCGGTTC GCGCGTAA GCGCGAAGG
 251 CCGCGCTTGC CTTGAACAAG GCGGGTTTGG CGTATTTTGA AGGCGGTTTT
 301 GAAAGCGCGG AACTAGAACG CTCACGCGTG TTGGTCAACA AAGAGGCGCG
 351 AGACAACCGG ACTTTGGCAT TGATGCTGGG CCGCGACGCC GCGCGACAAG
 401 TGGAAACAT CGAGCTGCGC GACCGTTATC TTGCGGAATC CGCCAAACTG
 451 CCGGAAAAAC AGCAGCTTTC CGGTTATCTT TTGTTGCGCG AATCGCGGCT
 501 GAACCGGCGC GATTACGAAG CCGCGGAAGC CAATCTTCAT GCGCGCGCGA
 551 AGATGAATGC CAACCTTAGC CGGCTCGTGC GTCTGCAACT TCGTTACGCT
 601 TTGACAGGGG GCGACGCGTT CGAGGTTCTG GCAGAAACCG AAAAATCTTC
 651 CAAGGCGGGC GCGTTGGGCA AATCGGAATG GGAACGGTAT CAAAATTGGG
 701 CATACGCGCG CGAGCTGGCG GATGCTGCGG ATGCGCGCGC TTTGAAAACC
 751 TGCGTGAAGC GGATTCCCGA CAGCCTCAAA AACGGGGAAT TGAGCGTATC
 801 GGTTCGGGAA AAGTAGGAAC GTTTGGGACT GTATCGCGAT GCGGTCAAAT
 851 GGGTCAAAAC GCATTATCCG CACAACCGCC CCGCGAGGCT TTGGAAGC
 901 TTTGCGAAA GCGTGGCGTT TTGCGCGAG CCGGACACG AGAAGCCAT
 951 CGATTTTGGC GATGCTTGGC TGAAGAACA GCCGATTAC GCGCTCTCG
 1001 TGATGTATCT CGGTCGGCTC GCGTACGGCC CAAACTTTG GGGCAAGGCA
 1051 AAAGGCTACC TTGAAGCGAG CATTCGATTA AAGCGAGTA TTTCCGCGCG
 1101 TTTGTTCTA GCAAGGTTTC TCGACGAAAT CGGAGACCG CAGAAGCGCG
 1151 AGGCGCAGC CAACTGGTT TTGGAAGCG TCTCCGATGA CGAACGTCAC
 1201 GCAGCGTTAG AGCAGCATAG CTGA

This corresponds to the amino acid sequence <SEQ ID 752; ORF100-1>:

1 MKTVVWIVVL FAAAVGLALA SGIYTGSDVYI VLQGTMLRIN LHAFVLGSLI
 51 AVVVVYFLFK FTIGVNLNPE KMQRFGSARK GRKALALANK AGLAYFEGRF
 101 EKAELEASRV LVNKEAGDNR TLALMLGAHA AGQMENTELR DRYLAEIAKL
 151 PEKQQLSRYL LLAESALNRR DYEAAEANLH AAKAMNANLT RLVLQLRYA
 201 FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRQLA DAADAAALKT
 251 CLKRI PDSLK NGELSVSVAE KYERLGLYD AVKWKQHYF HNRRLPELLEA
 301 FVESVRFLGE REQKAIIDFA DAWLKEQPDN ALLMYLGLR AYGRKLWGKA
 351 KGLEASIAL KESISARLV LAKVFEIDIEF QKAEQRNLV LEAVSDDERH
 401 AALEQHS*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF100 shows 93.5% identity over a 386aa overlap with an ORF (ORF100a) from strain A of *N.*

meningitidis:

		10	20	30	40	50	60
orf100.pep		MKTVVWIVLFAAAVGLALASGIYTGSDVYI	VLQGTMLRINLHAFVLGSLIAVVVYFLFK				
orf100a		MKTVVWIVLFAAAVGLALASGIYTGSDVYI	VLQGTMLRINLHAFVLGSLIAVVVYFLFK				
		10	20	30	40	50	60
		70	80	90	100	110	120
orf100.pep		FTIGVNLNPEKMQRFGSARKGXKXLLANL	KAGLAYFEGRFEKAELEASRVLVNKGDRNR				
orf100a		FTIGVNLNPEKMQRFGSARKGRKALALN	KAGLAYFEGRFEKAELEASRVLVNKEAGDNR				
		70	80	90	100	110	120
		130	140	150	160	170	180
orf100.pep		TLALMLGAHAAGQMENTELRDRYLAEIAKL	PEKQQLSRYLLLAESALNRRDYEAAEANLH				
orf100a		TLALMLGAHAAGQMENTELRDRYLAEIAKL	PEKQQLSRYLLLAESALNRRDYEAAEANLH				
		130	140	150	160	170	180
		190	200	210	220	230	240
orf100.pep		AAAKMNNANLRLVRLXIRYAFDRGDALQVL	AKTEKLSKAGALGKSEMERYQNWAYRQLA				
orf100a		AAAKMNNANLRLVRLQLRYAFDRGDALQVL	AKTEKLSKAGALGKSEMERYQNWAYRQLX				
		190	200	210	220	230	240
		250	260	270	280	290	300
orf100.pep		DAADAAALKTCLKRI PDSLKNGELSVSVAE	KYERLGLYDAVKWKQHYFKNRRPELLEA				
orf100a		DAADAAALKTCLKRI PDSLKNGELSVSVAE	KYERLGLYDAVKWKQHYFPHNRRPELLEA				
		250	260	270	280	290	300

		310	320	330	340	350	360
5	orf100.pep	FVESVRFGLGEREQQKAIDFADAWLKEQPDNALLMYLGRLAIFGRKLWGWKAGKYLEASIAL					
	orf100a	FVESVRFGLGERDQQKAIDFADAWLKEQPDNALLMYLGRLAYGRKLWGWKAGKYLEASIAL					
		310	320	330	340	350	360
		370	380				
10	orf100.pep	KPSISARLVLTKEVDEIGEPQKAEAH					
	orf100a	KPSISARLVLAKEVDEIGEPQKAEARQNLVLASVAEENRPSAETHX					
		370	380	390	400		

The complete length ORF100a nucleotide sequence <SEQ ID 753> is:

15	1	ATGAAAACGG	TAGTCTGGAT	TGTCGTCTCTG	TTTGCCGCGC	CNNTCGGGCT
	51	GCATTTGGCG	TGGGGCATTN	ACACCGGCGA	CGGTATATAT	GTACTCGGAC
	101	AGACCATGCT	CAGAATCAAC	CTGCACGCC	TTGTGTTAGG	TTCGCTGATT
	151	CGCGTGTGG	TGCGGTATT	CTGTTCACAA	TTCATCATCAT	CGCTACTCAA
	201	TANCCCGGAA	AAGATGCAGC	GTTTCGGTTC	GGCGCGTAA	GGCCGCAAGG
	251	CCGCGCTTGC	TTTGAACAA	GCGGGTTTGG	CGTATTTTGA	AGGGCGTTT
20	301	GAAAAGCGCG	AACCTGAAGC	CTCGCGGTA	TTGGGAAACA	AAGAGGCGGG
	351	GGATAACCGG	ACTTTGGCAT	TGATGTTGGG	CGCACATGCC	GCCGGGCGAG
	401	TGGAAACAT	CGAGCTGCGC	GACCGTTATC	TTGCGGAAAT	CGCCAACTG
	451	CGGAAAGAC	AGCAGCTTTC	CGGTTATCTT	TTGTTGGCGG	AATCGCGGCT
	501	GAACCGGCGC	GATTACGAAG	CGCGGGAAGC	CAATCTTCAT	GCGGCGGCGA
25	551	AGATGAATGC	CAACCTTAAG	CGCCTCGTGC	GTCTGCAACT	TCGTTACGCT
	601	TTGCACAGGG	GCGACCGGTT	GCAGGTTCTG	GCAAAACCCG	AAAAATTTTC
	651	CAAGGCGGGC	GCGTNGGCA	AATCGGAAAT	GGAAACGGTAT	CAAAATTTGG
	701	CATACCGCCG	CCAGCTGNCG	GATGCTGCGG	ATGCGCGCGC	TTTGAACACC
	751	TGCGCTGAAGC	GGATTCCCGA	CAGCCTCAAA	AACGGGGAAT	TSAGCGTATC
30	801	GGTTCGGGAA	AAGTACGAAC	GTTTGGGACT	GTATGCGCAT	CGCGTCAAT
	851	GGGTCAACCA	GCAATTATCG	CACACCGCG	GACCGCACT	TTTGAAGCN
	901	TTTGTGAAA	CGCTGCGCTT	TTTGGCGGAA	CCCGATCCAG	AGAAAGCCAT
	951	CGATTTTGCC	GATGCTTGGC	TGAAGAACA	CGCGATAAT	CGCGTCTCG
	1001	TGANGTATCT	CGGTGCGCTC	GCTACGGCC	CGAAACTTTG	GGGCAAGGCA
35	1051	AAAGGCTACC	TTGAAGCGAG	CATTGCATTA	AAGCGAGTA	TTTCCGCGG
	1101	TTTGTGTTCTG	GCAAAAGTTT	TTGACGAAC	CGGAGAACCG	CAGAAGCGG
	1151	AGGCGCAGCG	CAACTTGGTT	TTGGCAAGCG	TTGCGGAGGA	AAACCGNCT
	1201	TCCGCGGAAA	CCCATTTGA			

This encodes a protein having amino acid sequence <SEQ ID 754>:

40	1	MKTVVWIVVL	FAAXGLALA	SGIXTGDVYI	VLGQTMRLIN	LHAFVLGSLI
	51	AVVVVYFLEK	FIIGVLNXP	KMQRFSGARK	GRKAALALNK	AGLAYFEGRF
	101	EKAELASRV	LGNKEAGDNR	TLALMLGAHA	AGOMENIELR	DRYLAEIAKL
	151	PEKQQLSRYL	LLAESALNRR	DYEAEEANLH	AAAKMANLIT	RLVLQLRYA
	201	FDRGDALQVL	ARTEKSKAG	AXGKSEMERV	QNWAYRQLX	DAADAALIT
45	251	CLKRIPDSLK	NGELSVSAE	KYERLGLYAL	AVRWYRQHPY	HNRRPELLEA
	301	FVESVRFLE	RQQKAIDFA	DAWLKEQPDN	ALLLYLGLR	AYGRKLWGKA
	351	KGYLEASIAL	KPSISARLV	AKVFEDEGEP	QKAEARQNLV	LASVAEENRP
	401	SAETH*				

ORF100a and ORF100-1 show 95.1% identity in 406 aa overlap:

50		10	20	30	40	50	60
	orf100a.pep	MKTVVWIVVLF	FAAXGLALAS	GIXTGDVYI	VLGQTMRLIN	LHAFVLGSLI	AVVVVYFLEK
	orf100-1	MKTVVWIVVLF	FAAXVGLALAS	GIYTGQVYI	VLGQTMRLIN	LHAFVLGSLI	AVVVVYFLEK
55		10	20	30	40	50	60
	orf100a.pep	FIIGVLNPEK	MQRFSGARK	GRKAALALNK	AGLAYFEGR	FEKAELEASRV	LGNKEAGDNR
	orf100-1	FIIGVLNPEK	MQRFSGARK	GRKAALALNK	AGLAYFEGR	FEKAELEASRV	LGNKEAGDNR
60		70	80	90	100	110	120
	orf100a.pep	TLALMLGAHA	AGOMENIELR	DRYLAEIAKL	PEKQQLSRYL	LLAESALNRR	DYEAEEANLH
	orf100-1	TLALMLGAHA	AGOMENIELR	DRYLAEIAKL	PEKQQLSRYL	LLAESALNRR	DYEAEEANLH
65		130	140	150	160	170	180
	orf100a.pep	TLALMLGAHA	AGOMENIELR	DRYLAEIAKL	PEKQQLSRYL	LLAESALNRR	DYEAEEANLH
	orf100-1	TLALMLGAHA	AGOMENIELR	DRYLAEIAKL	PEKQQLSRYL	LLAESALNRR	DYEAEEANLH

		130	140	150	160	170	180
		190	200	210	220	230	240
5	orf100a.pep	AAAKNNANLTRLVRLQLRYAFDRGDALQVLAKTEKXSKAGAXGKSEMERYQNWAYRRQLX					
	orf100-1	AAAKNNANLTRLVRLQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQLA					
		190	200	210	220	230	240
		250	260	270	280	290	300
10	orf100a.pep	DAADAAALKTCLKRIPDSLKNGELSVSVAEKYERLGLYADAVKWKQHYHNNRRPELLEA					
	orf100-1	DAADAAALKTCLKRIPDSLKNGELSVSVAEKYERLGLYADAVKWKQHYHNNRRPELLEA					
		250	260	270	280	290	300
		310	320	330	340	350	360
15	orf100a.pep	FVESVRFIFGEREQQKAIQFADAWLKEQPDNALLLXLYLGRLAYGRKLWGKAKGYLEASTAL					
	orf100-1	FVESVRFIFGEREQQKAIQFADAWLKEQPDNALLLXLYLGRLAYGRKLWGKAKGYLEASTAL					
		310	320	330	340	350	360
20	orf100a.pep	KPSISARILVLAQVDETFGEFQKAEAOQNLVLASVAENRPSA-ETHX					
	orf100-1	KPSISARILVLAQVDETFGEFQKAEAOQNLVLASVADDERHAALEQHSX					
		370	380	390	400		
25		370	380	390	400		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF100 shows 93.3% identity over a 386 aa overlap with a predicted ORF (ORF100ng) from

N.gonorrhoeae:

30	orf100.pep	MKTVVWVIVLFAAAVGLALASGIYTGDIYIVLQTMRLNLHAFVLGSLIAVWVYFLFK	60
	orf100ng	MKTVVWVIVLFAAAVGLALASGIYTGDIYIVLQTMRLNLHAFVLGSLIAVWVYFLFK	60
		FIIGVLNIPEKMQRFGSARKGKXKXLAINKAGLAYFEGRFEKAELEASRVLNKVGDRNR	120
35	orf100.pep	FIIGVLNIPEKMQRFGSARKGKXKXLAINKAGLAYFEGRFEKAELEASRVLNKVGDRNR	120
	orf100ng	FIIGVLNIPENMRRSGSARKGKXKXLAINKAGLAYFEGRFEKAELEASRVLNKVGDRNR	120
		TLALMLXHAHAAGQMENIXXRDRYLAEIAKLPEKQQLSRYLLLAESALNRRDYEAEEANLH	180
40	orf100.pep	TLALMLXHAHAAGQMENIXXRDRYLAEIAKLPEKQQLSRYLLLAESALNRRDYEAEEANLH	180
	orf100ng	TLALMLXHAHAAGQMENIELDRYLAEIAKLPEKQQLSRYLLLAESALNRRDYEAEEANLH	180
		AAAKNNANLTRLVRLQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQLA	240
45	orf100.pep	AAAKNNANLTRLVRLQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQLA	240
	orf100ng	AAAKNNANLTRLVRLQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQLA	240
		DAADAAALKTCLKRIPDSLKNGELSVSVAEKYERLGLYADAVKWKQHYHNNRRPELLEA	300
	orf100ng	DAADAAALKTCLKRIPDSLKNGELSVSVAEKYERLGLYADAVKWKQHYHNNRRPELLEA	300
		FVESVRFIFGEREQQKAIQFADAWLKEQPDNALLLXLYLGRLAYGRKLWGKAKGYLEASTAL	360
50	orf100.pep	FVESVRFIFGEREQQKAIQFADAWLKEQPDNALLLXLYLGRLAYGRKLWGKAKGYLEASTAL	360
	orf100ng	FVESVRFIFGEREQQKAIQFADAWLKEQPDNALLLXLYLGRLAYGRKLWGKAKGYLEASTAL	360
		KPSISARILVLAQVDETFGEFQKAEAH	386
55	orf100.pep	KPSISARILVLAQVDETFGEFQKAEAH	386
	orf100ng	KPSIPARLVLAQVDETFGEFQKAEAOQNLVLASVAGENRPSAETR	405

The complete length ORF100ng nucleotide sequence <SEQ ID 755> is:

1	ATGAAACGG	TAGTCTGAT	TGTTGCTGT	TTTGGCGCG	CGCTGCGACT
51	GGCGCTGGCT	TCGGCATTT	ACACGGGGA	CGTGTATATC	GTACTCGGAC
101	AGACCATGCT	CAGAATCAAC	CTGACGGCT	TTGTGTTAGG	TTGCTGATT
151	CGCGCTGTGG	TGTGTGATTT	CTGTGTTAA	TTTCATCATCG	CGGTACTCAA
201	TATCCCCGAA	AATATGCGCG	GTTCCGGTC	GGCGCGGAA	GGCGCAAGG
251	CGCGCTGTGC	CTTGAATAG	CGGGGTTTG	CGTATTTCGA	AGGGCGTTT
301	GAAAGGCGG	AACTCGAAGC	CTCTCGAGT	TTGGGCAACA	AAGAGCGCG
351	AGACAACGG	ACTTTGGCAT	TGATGCTGG	CGCGACGGC	GCAGGACGAA
401	TGGAAATAT	CGAGCTGCGC	GACCGTTATC	TTGCGGAAT	CGCAAACTG

451 CCGGAAAAAC AGCAGCTTTC CCGCTATCTT CTGCTGGCGG AATCGGGGTT
 501 AAACCGCGGC GATTACGAGC CGCGCGAAGC CAATCTTCAT CGCGCGGCGA
 551 AGATGAATGC CAACCTTACG CGCCTCGTGC GTCTCGAATC TCGTTACGCC
 601 TTCGATCGGG GCGATGCGTT GAGGTTCTG GCAAAAaccG AAAAactTTT
 651 CAAGGCGGGC GCGTTGGGCA AATCGGAAAT GGAAACGATAT CAAAACTGGG
 701 CATACCGCGC CCAGATGGCG GATGCTGCGC ATGCGCGCGC TTGAAAAACC
 751 TGCTCGAAGC GGATTCGCCA CAGCCTCAAA AACGGGGAAT TGagcGTATC
 801 GGTTCGGGAA AAGTACGAA CTTTGGGACT GTATGCGAT CGCGTCAAT
 851 GGGTCARACA GCATTATCCG CACAAACCGC GCCCGAGCT TTGGAAGCC
 901 TTGTCGAAA GCGTGCCTT TTTGGCGAG CGCGAACACG AGAAAGCCAT
 951 CGATTTTGCC GATTCTGGC TGAAGAACA GCCCGATAAC CGCCTTCTGC
 1001 TGATGTATCT CGGCGGCGCT GCTACGGCC GCAAACTTTG GGGTAAGCGA
 1051 AAAGGCTACC TTGAAGCAG TATTCGACT AACCGCGAGT TTCGCGCGCG
 1101 TTTGGGTTG CCAAGGTTT TTGACGAAAC CGCACAGTCG CAAAAGCGCG
 1151 AAGCACAGCG CACTTGGTT TTGGCAAGCG TTGCGCGGGA AAAACGCGCT
 1201 TCCGCGGAAA CCGGTTGA

This encodes a protein having amino acid sequence <SEQ ID 756>:

1 MKTVVIVVL FAAAVGLALA SGIYTGdVYI VLGQTMRLIN LHAFLVLSLI
 51 AVVVWYFLFK FIIIGVNIPE NMRRSGSARK GRKAALALNK AGLAYFEGRF
 101 EKAELEASRV LGNKEAGDNR TLALMLGAHA AGQMNIELR DRYLAETAKL
 151 PEKQQLSRYL LLAESALNRR DYEAELANLH AAAMNNNLTL RLVRQLRYA
 201 FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQMA DAADAALKLT
 251 CLKRIPDSLK NGELSVSVAE KYERLGLYAD AVKWKQHYH HNNRPELLEA
 301 FVESVRFLEGE REQKKAIDFA DSWLKEQPDN ALLMYLGRLL AYGRKLWGKA
 351 KGYLEASIAL KPSPARLVL AKVFDETAQS QKAEAQRLNV LASVAGENRP
 401 SAETR*

ORF100ng and ORF100-1 show 95.3% identity in 402 aa overlap:

		10	20	30	40	50	60
30	orf100-1.pep	MKTVVIVVLFAAAVGLALASGIYTGdVYIVLGQTMRLINLHAFLVLSLIAVVVWYFLFK					
	orf100ng	MKTVVIVVLFAAAVGLALASGIYTGdVYIVLGQTMRLINLHAFLVLSLIAVVVWYFLFK					
		10	20	30	40	50	60
35	orf100-1.pep	FIIIGVNIPEKMQRFSGSARKGRKAALALNKAAGLAYFEGRFKAELEASRVLVNKEAGDNR	70	80	90	100	110
	orf100ng	FIIIGVNIPENMRRSGSARKGRKAALALNKAAGLAYFEGRFKAELEASRVLVNKEAGDNR	70	80	90	100	110
		130	140	150	160	170	180
40	orf100-1.pep	TLALMLGAHAAGQMNIELRDRLYLAEIAKLPEKQQLSRYL LLAESALNRRDYEAELANLH					
	orf100ng	TLALMLGAHAAGQMNIELRDRLYLAEIAKLPEKQQLSRYL LLAESALNRRDYEAELANLH					
		130	140	150	160	170	180
45	orf100-1.pep	AAAKMNNALTRLVRQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQLA	190	200	210	220	230
	orf100ng	AAAKMNNALTRLVRQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQMA	190	200	210	220	230
		250	260	270	280	290	300
50	orf100-1.pep	DAADAALKTCLKRIPDSLKNGELSVSVAEKYERLGLYADAVKWKQHYHNNRPELLEA					
	orf100ng	DAADAALKTCLKRIPDSLKNGELSVSVAEKYERLGLYADAVKWKQHYHNNRPELLEA					
		250	260	270	280	290	300
55	orf100-1.pep	FVESVRFLEGEREQKKAIDFADAWLKEQPDNALLMYLGRLAYGRKLWGKAKGYLEASIAL	310	320	330	340	350
	orf100ng	FVESVRFLEGEREQKKAIDFADSWLKEQPDNALLMYLGRLAYGRKLWGKAKGYLEASIAL	310	320	330	340	350
		370	380	390	400		
65	orf100-1.pep	KPSISARLVLAQVFDETAQSKAEAQRLNVLASVAGENRPSAETR					
	orf100n	KPSIPARLVLAQVFDETAQSKAEAQRLNVLASVAGENRPSAETR					

370 380 390 400

Based on this analysis, including the presence of a putative leader sequence, a putative transmembrane domain, and a RGD motif, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 90

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 757>

```

1  ATGATGTTTT  CTGGTTCAA  GCTGTTTAC  TTGTTTTTTG  TCATTTCGTG
10  51  GTTTCAGGG  CTGTTTTAC  TGCAGGAGAT  TTTCGTCAT  ATGGCGATGA
101  TTGATGTGCC  GCGCGGCAAT  CCGAGTATG  TGCCTCTGTC  GGCATGGCG
151  GTGCGGCTGT  ACGCTTTTAT  GTGCGCGTTG  GGCTTCGGCG  CGGTCTGTGT
201  CGGCGCGGCG  ATACCGTTTG  CCGCGCGGTG  GTGGGCGAGC  GGCTGGGTAC
251  ACGTCAAAC  GTGTTTGGGC  TTGATGCTCT  TGGCTTACCA  GTTGATTGCG
15  301  GGCCTGCTGC  TGCGCCGTTT  TCAGGATTAC  AGCAATGCTT  TTTCACACCG
351  CTGTACCGC  GTGTCAACG  AATCCCGT  GCTGCTGATG  GTTGGCGCGC
401  TGTATCTGGT  CGTGTTCAAA  CCGTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 758; ORF102>:

```

1  MMFSWFKLFH  LFFVISWFAG  LFYLPRIFFN  NAMIDVPRGN  PEYVRLSGMA
20  51  VRLYRFMSPL  GFGAVVFGAA  IPFAAGWWS  GWVHKLCGL  LMLLAYQLYC
101  GVLLRRFDY  SNAFSHRWYR  VFNEIPVLLM  VAALYLVVK  PF*

```

Further work revealed the complete nucleotide sequence <SEQ ID 759>:

```

1  ATGATGTTTT  CTGGTTCAA  GCTGTTTAC  TTGTTTTTTG  TCATTTCGTG
25  51  GTTTCAGGG  CTGTTTTAC  TGCAGGAGAT  TTTCGTCAT  ATGGCGATGA
101  TTGATGTGCC  GCGCGGCAAT  CCGAGTATG  TGCCTCTGTC  GGCATGGCG
151  GTGCGGCTGT  ACGCTTTTAT  GTGCGCGTTG  GGCTTCGGCG  CGGTCTGTGT
201  CGGCGCGGCG  ATACCGTTTG  CCGCGCGGTG  GTGGGCGAGC  GGCTGGGTAC
251  ACGTCAAAC  GTGTTTGGGC  TTGATGCTCT  TGGCTTACCA  GTTGATTGCG
301  GGCCTGCTGC  TGCGCCGTTT  TCAGGATTAC  AGCAATGCTT  TTTCACACCG
30  351  CTGTACCGC  GTGTCAACG  AATCCCGT  GCTGCTGATG  GTTGGCGCGC
401  TGTATCTGGT  CGTGTTCAAA  CCGTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 760; ORF102-1>:

```

1  MMFSWFKLFH  LFFVISWFAG  LFYLPRIFFN  NAMIDVPRGN  PEYVRLSGMA
35  51  VRLYRFMSPL  GFGAVVFGAA  IPFAAGWWS  GWVHKLCGL  LMLLAYQLYC
101  GVLLRRFDY  SNAFSHRWYR  VFNEIPVLLM  VAALYLVVK  PF*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with HP1484 hypothetical integral membrane protein of *H. pylori* (accession number AE000647)

ORF102 and HP1484 show 33% aa identity in 143aa overlap:

```

40  orf102  3  FSWFKLFHFFVISWVAGLFYLPRIFFNNAMIDVPRGNPEYVRLSGMAVRLYRFMSPLGF  62
    F W K EH+  VISW A LFYLP+EV  A  +  V++  +LY F++
HP1484  8  FLWVKAFHVIAVISWMAALFYLPRLFVYHAENAHKKEFVGTVQIEK--KLYSFIASPM  65

orf102  63  GAVVFGAAIPFAAG---WWGSGWVHKLCGLGMLLAYQLYCGVLLRRFDQYSNAFSHRWY  119
    G + + +  +  GW+H KL L ++LLAY YC +R + + R+Y
45  HP1484  66  GFTLTITGLIMLLIEPTLFKSGWHLHAKLALVLLLAYHFYCKKCMRELEKDPTRRNARY  125

orf102  120  RVFNEIPXXXXXXXXXXXXFKPF  142
    RVFNE  F  KPF
HP1484  126  RVFNEAPTILMILVILVVKPF  148

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF102 shows 99.3% identity over a 142aa overlap with an ORF (ORF102a) from strain A of *N.meningitidis*:

5	orf102.pep	10	20	30	40	50	60
	MMFSWFKLHFLFFVISWVAGLFYLPRIFFNMAMIDVPRGNPEYVRLSGMAVRLYRFMSPL						
	orf102a	10	20	30	40	50	60
	MMFSWFKLHFLFFVISWVAGLFYLPRIFFNMAMIDVPRGNPEYVRLSGMAVRLYRFMSPL						
10	orf102.pep	70	80	90	100	110	120
	GFGAVVFGAAIPFAAGWWSGWVHVKLCGLMLLAYQLYCGVLLRRFPQDYSNAFSHRWYR						
	orf102a	70	80	90	100	110	120
	GFGAVVFGAAIPFAAGWWSGWVHVKLCGLMLLAYQLYCGVLLRRFPQDYSNAFSHRWYR						
15	orf102.pep	130	140				
	VFNEIPVLLMVAALYLVVFKPF						
	orf102a	130	140				
	VFNEIPVLLMVAALYLVVFKPF						

The complete length ORF102a nucleotide sequence <SEQ ID 761> is:

1	ATGATGTTTT	CTTGTTTCAA	GCTGTTTCAC	TTGTTTTTGG	TCATTTTCGTG
51	GTTTGCAGGG	CTGTTTTCAC	TGCCGAGGAT	TTTCGTCAAT	ATGCGCATGA
25	TTGATGTGCC	GCGCGCAAT	CCCGAGTATG	TGCGTCTGTC	GGSCATGGCG
151	GTCGCGCTGT	ACCGTTTAT	GTCGCGCTTG	GGCTTCGCGC	CGGTCTGTGT
201	CGCGCGCGCG	ATACCGTTTG	CCCGCGCTG	GTCGCGGAC	GGCTCGGAC
251	ACGTCAACT	CTGTTTGGC	TGATGCTCT	TGCTTACCA	GTTTATTGG
301	GGCGTGTGTC	TGCGCGCTT	TCAGGATTAC	ACCAATGCTT	TTTCACACCG
351	CTGTACCGC	GTGTTCAACG	AAATCCCGT	GCTGCTGATG	GTTGCGCGCG
401	TGATCTGGT	CGTGTCAA	CGTGTGTA		

This encodes a protein having amino acid sequence <SEQ ID 762>:

1	MMFSWFKLH	LFFVISWVAG	LFYLPRIFFN	MAMIDVPRGN	PEYVRLSGMA
51	VRLYRFMSPL	GFGAVVFGAA	IPFAAGWWSG	GWVHVKLCGL	MLLAYQLYCG
35	GVLLRRFPQDY	SNAFSHRWYR	VFNEIPVLLM	VAALYLVVFK	PF*

ORF102a and ORF102-1 show complete identity in 142 aa overlap:

	orf102a.pep	10	20	30	40	50	60
	MMFSWFKLHFLFFVISWVAGLFYLPRIFFNMAMIDVPRGNPEYVRLSGMAVRLYRFMSPL						
40	orf102-1	10	20	30	40	50	60
	MMFSWFKLHFLFFVISWVAGLFYLPRIFFNMAMIDVPRGNPEYVRLSGMAVRLYRFMSPL						
	orf102a.pep	70	80	90	100	110	120
	GFGAVVFGAAIPFAAGWWSGWVHVKLCGLMLLAYQLYCGVLLRRFPQDYSNAFSHRWYR						
45	orf102-1	70	80	90	100	110	120
	GFGAVVFGAAIPFAAGWWSGWVHVKLCGLMLLAYQLYCGVLLRRFPQDYSNAFSHRWYR						
	orf102a.pep	130	140				
	VFNEIPVLLMVAALYLVVFKPF						
	orf102-1	130	140				
	VFNEIPVLLMVAALYLVVFKPF						

55 Homology with a predicted ORF from *N.gonorrhoeae*

ORF102 shows 97.9% identity over a 142 aa overlap with a predicted ORF (ORF102ng) from *N.gonorrhoeae*:

-426-

```

    orf102.pep      MMFSWFKLFHLFFVISWFAGLFYLPRI FVNAMIDVPRGNPEYVRLSGMAVRLYRFMSPL 60
                   |||
    orf102.ng      MMFSWFKLFHLFFVISWFAGLFYLPRI FVNAMIDVPRGNPEYVRLSGMAVRLYRFMSPL 60
                   |||
5    orf102.pep      GFGAVVFGAAIPFAAGRWGSGWVHVKLCGLMLLAYQLYCGVLLRRFPQDYSNAFSHRWYR 120
                   |||
    orf102.ng      GFGAVVFGAAIPFAAGRWGSGWVHVKLCGLMLLAYQLYCGVLLRRFPQDYSNAFSHRWYR 120
                   |||
10   orf102.pep      VFNEIPVLLMVAALYLVVFKPF 142
                   |||
    orf102.ng      VFNEIPVLLMVAALYLVVFKPF 142
                   |||

```

The complete length ORF102ng nucleotide sequence <SEQ ID 763> is:

```

1   ATGATGTTT CTGTTTCAA GCTGTTTAC TTGTTTTTG TCAFTTCGTG
5   GTTTGCAGGG CTGTTTACC TGGCGAGGAT TTCTCTCAAT ATGCGCATGA
15  TTGATCGCCC CGCGCGCAT CCAGAGTATG TCGCGCTGTC GGGGATGGCG
151 GTGCGGTGT ACCGTTTAT GTCGCTTTG GGTTCGGCG CGGTCTGTTT
201 CGCGCGGGCG ATACGTTTG CGCGggcgcg GTGGGGCagc ggtcggGTTT
251 ACGTCAAAC TGTGTTGGG TGTGCTCTT TGGCTATCA GTTGATTGCG
301 GCGGTGCTGC TGCGCCGTT TCAGGATTAC AGCAATGCT TTTCACACCG
20  351 CTGTGACGCG GTGTTCAAC aATCCCGT GCTGCTGAT GTTGCCGCGC
401 TGTATCTGT CGTGTCAA CCGTTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 764>:

```

1   MMFSWFKLFH LFFVISWFA LFLYLPRI FVN MAMIDAPRN PEYVRLSGMA
25  51 VRLYRFMSPL GFGAVVFGAA IPFAAGRWGS GWVHVKLCGL LMLLAYQLYC
101 GVLLRRFPQDY SNAFSHRWYR VFNEIPVLLM VAALYLVVFK PF*

```

ORF102ng and ORF102-1 show 98.6% identity in 142 aa overlap:

```

               10      20      30      40      50      60
    orf102-1.pep MMFSWFKLFHLFFVISWFAGLFYLPRI FVNAMIDVPRGNPEYVRLSGMAVRLYRFMSPL
30   orf102.ng   MMFSWFKLFHLFFVISWFAGLFYLPRI FVNAMIDVPRGNPEYVRLSGMAVRLYRFMSPL
               10      20      30      40      50      60
    orf102-1.pep GFGAVVFGAAIPFAAGRWGSGWVHVKLCGLMLLAYQLYCGVLLRRFPQDYSNAFSHRWYR
35   orf102.ng   GFGAVVFGAAIPFAAGRWGSGWVHVKLCGLMLLAYQLYCGVLLRRFPQDYSNAFSHRWYR
               70      80      90     100     110     120
    orf102-1.pep VFNEIPVLLMVAALYLVVFKPF
40   orf102.ng   VFNEIPVLLMVAALYLVVFKPF
               130     140
    orf102-1.pep
    orf102.ng
               130     140

```

In addition, ORF102ng shows significant homology to a membrane protein from *H. pylori*:

```

45  gi|2314656 (AE000647) conserved hypothetical integral membrane protein
    [Helicobacter pylori] Length = 148
    Score = 79.2 bits (192), Expect = 1e-14
    Identities = 50/147 (34%), Positives = 68/147 (46%), Gaps = 13/147 (8%)
50  Query: 3   FSWFKLFHLFFVISWFAGLFYLPRI FVNAMIDVPRGNPEYVRLSGMAVRLYRFMSPLGF 62
    F W K FH+ VISW A LFLYLPRI FV A + V++ +LY F++
    Sbjct: 8   PLWVKAPHVIAVISMAALFYLPRLFVYHAENAHKKEFVGQVQIQEK--KLYSFIASPM 65
55  Query: 63   GAVVFGAAIP-----FAAGRWGSGWVHVKLCGLMLLAYQLYCGVLLRRFPQDYSNAFS 115
    G + + F +G GW+H KL L ++LLAY YC +R + +
    Sbjct: 66   GFTLTIGIIMLLIEPTLFKSG---GWLHAKLALVLLLAYHFYCKKCMRELEKDPTRRN 121
    Query: 116 HRWYRVFNEIPXXXXXXXKFKPF 142
    R+YRVFNE P KPF
60  Sbjct: 122 ARFYRVFNEAPTILMILVILVVVFKPF 148

```

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 91

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 765>:

```

5      1  ATGGCAAAAA  TGATGAATG  GGCGGCTGTT  GCGCGGCTCG  CGCGCGCAGC
51     51  GGTTTGGGGG  GGAATGGCTTT  AACTGAAGCC  CGAGCCGCAC  GTGCTTGATA
101    101  TTACGGAAC  GGTCTAGCGCC  GGC // .....
//... ATTTCGTTA  CGATTTTGTG  CGAACCGGAT  ACGCGATTA  AGCGAAGCT
51     51  CGACAGCGTC  GACCCCGGGC  TGACCAAGAT  GTCGTCGGC  GGTTACAACA
101    101  GCAGTACGGA  TACGGCTTCC  AATCGGCTCT  ACTATTATGC  CCGTTCGTTT
151    151  GTGCCGAATC  CGGACGGCAA  ACTCGCCACG  GGGATGACGA  CGCAGAAATC
201    201  GGTTGAAATC  GACGGCGTGA  AAAATGTGCT  GATTATTCCG  TCGCTGACCG
251    251  TGA AAAATCG  CGCGCGCAG  GCGTTTGTGC  GCGTGTGGG  TCGGACGCG
301    301  AAGGCGGGCG  ACGCGAATAT  CCGGACCGGT  ATGAGAGACA  GTATGAATAC
151    351  CGAAGTAAAA  AGCGGTTTGA  AAGAGGGGGA  CAAAGTGCT  ATCTCCGAAA
401    401  TAAACGCGCG  CGAGCAACAG  GAAAGCGGCG  AACGCGCCT  AGCGCGCCCG
451    451  CCGCGCCGAT  AA

```

This corresponds to the amino acid sequence <SEQ ID 766; ORF85>:

```

20      1  MAKMKWAAV  AAVAAAAVWG  GWS.LKPEPH  VLDITETVRR  G.....
51     51  .....
101    101  .....
151    151  .....
201    201  ..... I SFTILSEPT
251    251  PIKAKLDSVD  PGLTMMSSGG  YNSDSTASN  AVVYIARFV  PNPDKLATG
25     301  MTTQNTVEID  GVKNVLIIPS  LTVKNRGGKA  FVRVLADGK  AAEREIRTMG
351    351  RDSMNTVEKS  GLKEGDKVVI  SEITAAEQE  SGERALGGPP  RR*

```

Further work revealed the further partial nucleotide sequence <SEQ ID 767>:

```

30      1  ..GTATCGGTCG  GCGCGCAGGC  ATCGGGCGAG  ATTAAGATAC  TTTATGTCAA
51     51  ACTCGGGCAA  CAGGTTAAAA  AGGGCGATTT  GATTGCGGAA  ATCAATTGCA
101    101  CCTCGCAGAC  CAATACGCTC  AATACGGAAA  AATCCAAGTT  GGAAACGTAT
151    151  CAGGCGAAGC  TGGTGTCCGC  ACAGATTGCA  TTGGGACGCG  CGGAGAAGAA
201    201  ATATAAGCGT  CAGCGCGGCT  TATGGAAGGA  AAGCGGACT  TCCAAAGAGG
251    251  ATTTGGAAG  CGCGCAGGAT  GCGTTTGCG  CGCCAAAGC  CAATGTTGCC
301    301  GAGCTGAAGC  CTTTAAATCAG  ACAGAGCAAA  ATTTCCATCA  ATACCGCCGA
35     351  GTCGGAAATTG  GGCTACACGC  GCATTACCGC  AACGATGGAC  GGCACGGTGG
401    401  TGCGGATTTCT  CGTGGAAAGG  GGGCGAGCTG  TGAACGCGCG  CGAGTCTACG
451    451  CGCAGGATTG  TCCAATTGGC  GAATCTGGAT  ATGATGTTGA  ACAAAATGCA
501    501  GATTTCGCGG  GCGCATATTA  CCAAGTGA  GCGGCGGCG  GATATTTCGT
551    551  TTACGATTTT  GTCCGACCGC  GATAGCCGGA  TTAAGCGTGA  GTCGACAGC
40     601  GTCGACCCCG  GCGCTGACCAC  GATGTCGTG  GCGGCTGACA  ACAGACGTAC
651    651  GGATACGCGT  TCCAAATGCG  TCTACTATTA  TGCOCGTTCG  TTTGTGCCGA
701    701  ATCCGCGACG  CAAACTCGCC  ACGGGGATGA  CGACGAGAA  TACGGTTGAA
751    751  ATCGACGGCG  TGA AAAATGT  GCTGATTATT  CCGTCGCTGA  CCGTGAAAAA
801    801  TCGCGGCGGC  AAGGCGTTTG  TCGCGGTGTT  GGGTCGGGAC  GGCRAAGCGG
45     851  CGGAACGCGA  AATCGGAC  GGTATGAGAG  ACAGATGAA  TACCGAAGTA
901    901  AAAAGCGGCT  TGAAGAGG  GGACAAAGTG  GTCATCTCCG  AAATAACCCG
951    951  CGCCGAGCAA  CAGGAAAGCG  CGGAACGCG  CCTAGCGCG  CCGCGCGCC
1001   1001  GATAA

```

This corresponds to the amino acid sequence <SEQ ID 768; ORF85-1>:

```

50      1  ..VSVGAQASG  IKILYVKLGO  QVKKGDLIAE  INSTQNTNL  NTEKSKLETY
51     51  QRKLVSQAIA  LGSAEKKYKR  QAALWKENAT  SKEDLESAQD  AFAPAKANVA
101    101  ELKALIRQSK  ISINTAESEL  GYTRITATMD  GTVVAILVEE  GQTVNAAQST
151    151  PTIVOLANLD  MMLNRMQIAE  GDITKVKAGO  DISFTILSEP  DTFPIKALDS
201    201  VDPGLTMMSS  GGYNSDTDA  SNAVYYIARS  FVNPDKGLA  TGMTQNTVE
55     251  IDGVKNVLI  PSTLVNRRGG  KAFVRLGAD  GKAAREIRLT  GMRDSMNTVE
301    301  KSLGKEGDKV  VISEITAAEQ  QESGERALGG  PPRR*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF85 shows 87.8% identity over a 41aa overlap and 99.3% identity over a 153aa overlap with an ORF (ORF85a) from strain A of *N. meningitidis*:

		10	20	30	40	
5	orf85.pep	MAKMMKWA	AAVAAA	AVVGGWS	-LKPEPHVLDITETVRRG	
	orf85a	MAKMMKWA	AAVAAA	AVVGGWSY	LKPEPQAAYITETVRRGDISRTVSATGEISPSNLVS	
		10	20	30	40	50 60
				//		
10	orf85.pep	ISFTILSE	PDTPPIKAKLDSVDPGLTMTSSG	
	orf85a	TIVQLANL	DMMLNKM	QIAEGDIT	KVKAGQDISFTILSEPDTPPIKAKLDSVDPGLTMTSSG	
15		210	220	230	240	250 260
	orf85.pep	110	120	130	140	150 160
	orf85a	GYNSSTD	TASNAV	YVYARS	FFVNFDPGK	LATGMTTQNTVEIDGVKNVLIIPSLTVKNRGGK
20		270	280	290	300	310 320
	orf85a	GYNSSTD	TASNAV	YVYARS	FFVNFDPGK	LATGMTTQNTVEIDGVKNVLIIPSLTVKNRGGK
25		170	180	190	200	210 220
	orf85.pep	AFVVRVL	GADGKA	AREIR	TGMGRDSM	NTEVSKLKEGDKVVI
	orf85a	AFVVRVL	GADGKA	AREIR	TGMGRDSM	NTEVSKLKEGDKVVI
30		330	340	350	360	370 380
	orf85.pep	230	PRRX			
	orf85a	390	PRRX			

The complete length ORF85a nucleotide sequence <SEQ ID 769> is:

35	1	ATGGCAAAA	TGATGAAATG	GGCGGCTGTT	CGCGGGTGC	CGGCGGCACG
	51	GGTTTGGGGC	GGATGGTCTT	ATCTGAAGCC	CGAGCCGACG	GCTGCTTATA
	101	TTACGGAAAC	GGTCAGGGCG	GGCGACATCA	CGCGGACGGT	TTCTGCAACA
	151	GGGGAGATT	GGCGGTCCAA	CCTGGTATCG	GTGCGCGCGC	AGGCATCGGG
	201	CGAGATTAA	AAACTTTATG	TCAAATCGCG	GCAACAGGTT	AAAAGGGCG
40	251	ATTTGATTGC	GGAATCAAT	TCGACCTCGC	AGACCAATAC	GCTCAATACG
	301	GAAAATCCA	AATTGGAAC	GTATCAGGCG	AAGCTGGTGT	CGGCACAGAT
	351	TGCATTGGCG	AGCGCGGAGA	AGAAATATA	GGTCAAGCGG	CGGTTGTGGA
	401	AGGATGATCG	GACCGCTAAA	GAGATTTGG	AAAGCGGCAC	GGATGGCGCT
	451	CGCGCGCCCA	AGCCCAATGT	TGCCGACTGT	AGGCTCTTAA	TGACACAGAG
	501	CARATTTCCT	ATCAATACCG	CCGACTCGGA	ATTGGGCTAC	ACCGCATTA
45	551	CCGCAACGAT	GGACGGCAGC	GTGGTGGCGA	TTCTGTGGGA	AGAGGGGACG
	601	ACTGTGAAGC	CGGCGCAGTC	TACGCCGAGC	ATTGTCCAAT	TGGCGAATCT
	651	GGATATGATG	TTGAACAAA	TGCAGATTGC	CGAGGGCGAT	ATTACCAAGG
	701	TGAAGCGGGG	GCAGGATATT	TGCTTTACGA	TTTTGTCCGA	ACCGGATACG
50	751	CCGATTAAAG	CGAAGCTCGA	CAGCGTCGAC	CCCGGGCTGA	CCACGATGTC
	801	GTCCGGGGCG	TACAACAGCA	GTACCGGATC	GGCTTCCAAT	CGGCTCTACT
	851	ATTATGCCCG	TTGGTTTGTG	CCGAATCCGG	ACGGCAAACT	CGCCACGGGG
	901	ATGACGACGC	AGAAATACGGT	TGAAATCGAC	GGTGTGAA	ATGTGCTGAT
	951	TATTCCTGCG	CTGACCGTGA	AAAAATCGCG	CGCGACGGCG	TTTGTGCGCG
	1001	TGTTGGGTGC	AGACGGCAGG	CGGGCGGAAC	GCGAAATTCG	GACCGGTATG
55	1051	AGAGACAGTA	TGAATACCGA	AGTAAARAAG	GGTGTGAAG	AGGGGGACAA
	1101	AGTGGTCATC	TCGGAATAAA	CCCGCGCCGA	GCAGCAGGAA	AGGGGCGAAC
	1151	GGCGCCCTAGG	CGGCCCGCGC	CGCCGATAA		

This encodes a protein having amino acid sequence <SEQ ID 770>:

60	1	MAKMMKWA	AAVAAA	AVVGGWS	YLYKPEPQ	AAAYITETVRR	GDISRTVSAT
	51	GEISPSNLVS	VGAQASGQIK	KLTVKLQGVV	KRGDLIAEIN	STSPNTLNT	
	101	EKSKLSTPQA	KLVSQAIALG	SAKKYKRRQA	ALWKDDIAAK	EDLESADRL	
	151	AAIAHVAEL	KALIRQSKIS	INTAESELY	TRITATMDGT	VVAILEEQQ	
	201	TVNAAGSFFT	TIVQLANLDM	LNKMQIAEGD	ITKVKAGQDI	FTILSEPEPT	
	251	PIKAKLDSVD	PGLTMTSSGG	YNSSTDASN	AVVYARSEV	PNPDGKLATG	
65	301	MTQNTVEID	GVKNVLIIPS	LTVKNRGGRA	FVRVLGADGK	AAEREIRTM	

351 RDSMNTVEVKS GLKEGDKVVI SEITAAEQQE SGERALGGPP RR*

ORF85a and ORF85-1 show 98.2% identity in 334 aa overlap:

		30	40	50	60	70	80	
5	orf85a.pep	PQAAIYITETVRRGDISRTVSATGEISPSNLVS	GAQASGQIKKLYVKLGQQVKKGD	LIAE				
	orf85-1		VSVGAQASGQIKKLYVKLGQQVKKGD	LIAE				
			10	20	30			
10	orf85a.pep	90	100	110	120	130	140	
	orf85a.pep	INSTSQTNTLNTTEKSKLETYQAKLVSAQIALGSAEKYKQAAALWKDDATAKEDLESAQD						
	orf85-1	INSTSQTNTLNTTEKSKLETYQAKLVSAQIALGSAEKYKQAAALWKENATSKEDLESAQD						
		40	50	60	70	80	90	
15	orf85a.pep	150	160	170	180	190	200	
	orf85a.pep	ALAAAKANVAELKALIROSKISINTAESLGYYTRITATMDGTVVAILVEEQGVNAAQST						
	orf85-1	AFAAAKANVAELKALIROSKISINTAESLGYYTRITATMDGTVVAILVEEQGVNAAQST						
		100	110	120	130	140	150	
20	orf85a.pep	210	220	230	240	250	260	
	orf85a.pep	PTIVQLANLDMMLNKMQIAEGDITVKAGQDISFTILSEPDTPIKAKLDSVDPLGTTMSS						
	orf85-1	PTIVQLANLDMMLNKMQIAEGDITVKAGQDISFTILSEPDTPIKAKLDSVDPLGTTMSS						
		160	170	180	190	200	210	
25	orf85a.pep	270	280	290	300	310	320	
	orf85a.pep	GGYNSSTDTASNAVYYARSEVFNPDGKLTGMMTQNTVEIDGVKNVLIIPSLTVKNRGG						
	orf85-1	GGYNSSTDTASNAVYYARSEVFNPDGKLTGMMTQNTVEIDGVKNVLIIPSLTVKNRGG						
		220	230	240	250	260	270	
30	orf85a.pep	330	340	350	360	370	380	
	orf85a.pep	RAFVVRVLGADGKAAEREIRTMGRDSMNTVEVKSGLKEGDKVVISEITAAEQQESGERALGG						
	orf85-1	KAFVVRVLGADGKAAEREIRTMGRDSMNTVEVKSGLKEGDKVVISEITAAEQQESGERALGG						
		280	290	300	310	320	330	
40	orf85a.pep	390						
	orf85a.pep	PFRRX						
	orf85-1	PFRRX						

Figure 19D shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF85a.

Homology with a predicted ORF from *N.gonorrhoeae*

45 ORF85 shows a high degree of identity with a predicted ORF (ORF85ng) from *N.gonorrhoeae*:

	ORF85	1	MAKMMKWA AVAAVAAAAVWGGWS.LKPEPHVLDTITVRRG.....	40
	ORF85ng	1	MAKMMKWA AVAAVAAAAVWGGWSYLKPEPQAAIYITEAVRRGDISRTVSAT	50
50	ORF85	ISFTILSEPD	250
	ORF85ng	201	TVNAAQSTPTIVQLANLDMMLNKMQIAEGDITVKAGQDISFTILSEPD	250
55	ORF85	251	PIKAKLDSVDPLGTTMSSGGYNSSTDTASNAVYYARSEVFNPDGKLATG	300
	ORF85ng	251	PIKAKLDSVDPLGTTMSSGGYNSSTDTASNAVYYARSEVFNPDGKLATG	300
60	ORF85	301	MTTQNTVEIDGVKNVLIIPSLTVKNRGGKAFVVRVLGADGKAAEREIRTM	350
	ORF85ng	301	MTTQNTVEIDGVKNVLIIPSLTVKNRGGKAFVVRVLGADGKAAEREIRTM	350
	ORF85	152	RDSMNTVEVKSGLKEGDKVVISEITAAEQQESGERALGGPFRX	393
65	ORF85ng	351	KDSMNTVEVKSGLKEGDKVVISEITAAEQQESGERALGGPFRX	393

The complete length ORF85ng nucleotide sequence <SEQ ID 771> is:

```

1   ATGGCAAAAA   TGATGAAATG   GGC GGCTGTT   GCGGCGGTCG   CGGCGGCaac
51  GTTTGGGGGC   GGATGGTCTT   ATCTGAAGCC   CGAACCCGAC   GCTGCTTATA
101 TTA CGGAaac   ggtCAGGGCG   GCGGATATCA   CGCGACGGGT   TTCGCGCAGC
151 GgcgAGATTT   CGCGCTCCAA   CTGCTATCG   GTCGCGCGGT   AGCTCTGAGC
201 GCAGATTA   AAGCTTTATG   TCAACTCGG   GCACACGGTC   AAAAAGGGCG
251 ATTGATTGC   GGAATCAAT   TCGACCACG   AGACCAACAC   GATCGATATG
301 GAAANAATCCA   AATTGCAAC   GTATCAGCG   AAGCTGGTGT   CGGCACAGAT
351 TGCATTGGCG   AGCCGCGAGA   AGAATATAA   CGCTCAGGG   GCGTTGTGGA
401 AGGATGATGC   GACCTCTAA   GARGATTGG   AAGCGCGGCA   GGATGCGCTT
451 GCGCGCGCCA   AAGCCAATGT   TGCCGAGTTG   AAGGCTTTAA   TCACACAGAC
501 CAAAATTTC   ATCAATACCG   CCGAGTCGGA   TTTGGGCTAC   ACGCGCATTA
551 CCGCGACGAT   GGAACGCAAG   GTGGTGCGGA   TTCCCGTGA   AGAGGGGCG
601 ACTGTGAACG   CGGCGCAGTC   TAOCGCCAGC   AITGTCCAAT   TGGCGAATCT
651 GGATATGATG   TTGAACAAA   TGCAGATTGC   CGAGGGCGAT   ATTACCAAGC
701 TGAAGGCGGG   CGAGGATATT   TCGTTTACGA   TTTTGTCCGA   ACCGGATAGC
751 CGGATTAAAG   CGAAGCTCGA   CAGCGTCGAC   CCGGGCTGA   CCACGATGTC
801 GTCGGGCGGC   TACAACAGCA   GTAOCGATAC   GCGTTCCAAT   GCGTCTATT
851 ATTATGCGCC   TCGTCTTGTG   CCGAATCCGG   ACGGCAAACT   CGCCACGGGG
901 ATGACGACGC   AGAATACGGT   TGAATCGAC   GGTGTGAAA   ATGTGTGGCT
951 TATTCCGTCG   CTGACCGTGA   AATTCGCG   CGCGACGGC   TTCGATCGCG
1001 TGTTGGGTGC   GGAACGCAAG   CGAGTGAAC   CGAATTCG   GACCGGATG
1051 AAGACACTA   TGAATACGGA   AGTGAAGAG   GGTGTGAAG   AGGGGACAA
1101 ACTGTCATC   TCGAATATA   CCGCGCGCA   CGACGAGGA   AGCGCGAAC
1151 GCGCCCTAGG   CGGCGCGCG   CGCGGATAA

```

This encodes a protein having amino acid sequence <SEQ ID 772>:

```

1   MAKMKAAV   AAVAAAAMV   GWSYLKPEPQ   AAYITEAVR   GDISRTVSAT
51  GRISPSNLVS   VQAQASGQIK   KLYVKLQOQV   KKGDLIAEIN   STQTNTIMD
101 EKSLETYQA   KLVSAQIALG   SAEKKYKRA   ALWKDDATSK   EDLESQDAL
151 AAAKANVAEL   KALIRQSKIS   INTAESDLGY   TRITATMDGT   VVALPVEEGQ
201 TVNAAQSPT   IVQLANLDM   LNMKQIAEGD   ITKVKAGQDI   SFTILSEPT
251 PIKAKLDSVD   PGLTMTSSGG   VNSSTDASN   AVYYARSEFV   PNPDKLATG
301 MTTQNTVEID   GVKNVLLIPS   LTVKNRGSKA   FVRVLGADGK   AVEREIRTGM
351 KDSMTNEVKS   LKREGDKVVI   SEITAAEQE   SGERALGGPP   RR*

35  ORF85ng and ORF85-1 show 96.1% identity in 334 aa overlap:

      30      40      50      60      70      80
orf85ng   PQAA YITETVRRGDISRTVSATGEISPSNLVS VQAQASGQIKKLYVKLQOQVKKGDLIAE
      100      110      120      130      140
orf85-1   VSVGAQASQIKILYVKLQOQVKKGDLIAE

      90      100      110      120      130      140
orf85ng   INSTTQNTNTIMEKSKLETYQAKLVSAQIALGSAEKKYKRAALWKDDATSKEDLESAQD
      150      160      170      180      190      200
orf85-1   INSTSQNTNTLITEKSKLETYQAKLVSAQIALGSAEKKYKRAALWKENATSKEDLESAQD

      150      160      170      180      190      200
orf85ng   ALAAAKANVAELKALIRQSKISINTAESDLGYTRITATMDGTVVAILPVEEGQTVNAAQST
      210      220      230      240      250      260
orf85-1   AFAAAKANVAELKALIRQSKISINTAESDLGYTRITATMDGTVVAILVEEGQTVNAAQST

      210      220      230      240      250      260
orf85ng   PTIVQLANLDM LNMKQIAEGDITKVKAGQDISFTILSEPTPIKAKLDSVD PGLTMTSS
      270      280      290      300      310      320
orf85-1   PTIVQLANLDM LNMKQIAEGDITKVKAGQDISFTILSEPTPIKAKLDSVD PGLTMTSS

      270      280      290      300      310      320
orf85ng   GGYNSSTDASN AVYYARSEFV PNPDKLATGMTTQNTVEIDGKVNLLIPSLTVKNRGG
      330      340      350      360      370      380
orf85-1   GGYNSSTDASN AVYYARSEFV PNPDKLATGMTTQNTVEIDGKVNLLIPSLTVKNRGG

```

-431-

```

orf85ng      KAFVRVLGADGKAVEREIRTGMKDSMNTVEVSKLKEGDKVVISSEITAAEQQESGERALGG
|||||
orf85-1      KAFVRVLGADGKAAREIRTGMKDSMNTVEVSKLKEGDKVVISSEITAAEQQESGERALGG
                280      290      300      310      320      330

orf85ng      390
P PRRX
|
orf85-1      P PRRX

```

10 In addition, ORF85ng shows significant homology to an *E.coli* membrane fusion protein:

```

gi|1787104 (AE000189) c380; 27% identical (27 gaps) to 332 residues from
membrane fusion protein precursor, MTRC_NEIGO SW: P43505 (412 aa) [Escherichia
coli] Length = 380
Score = 193 bits (485), Expect = 2e-48
Identities = 120/345 (34%), Positives = 182/345 (51%), Gaps = 13/345 (3%)

Query: 29 PQAAITVTVRGDISRTVSATGEISPSNLVSQAASGQIKKLVKLGQQVFKGDLIAE 88
      F Y T VR GD++ +V ATG++ V VQAQ SQ+K L V +G +VKR L+
Sbjct: 41 EVPTYTQLTLIVREGDLQQSVLATGKLDALAKVDVGAQVSGQLKTLVAIGDKVKDQLLGV 100

Query: 89 INSTTQTNTIDMEKSKLETYQAKLVSAQIALGSAEKYKRAALWKDDATSKSEXXXXXX 148
      I+ N I ++ L +A+ A+ L A Y RQ L + A S++
Sbjct: 101 IDPEQAEHQIKEVEATLMELRAQRQCAEAELKLARVTVTSRQRLAQTKAVSQQDLDTAAT 160

Query: 149 XXXXXXXXXXXXXXXXIRQSKISINTAESDLGYTRITATMDGTVAIPVEGQTVNAAQST 208
      I++++ S++TA+++L YTRI A M G V I +GGTV AAQ
Sbjct: 161 EMAVKQAQIGTIDAQIKRNQASLDATAKTLNLDYTRIVAFMAGEVTQITTLQGQTVIAAQA 220

Query: 209 FTIVQLANLDMMLINKMQIAEGDITKVKAGQDISFTILSEFPTPIKAKLDSVDPGLTTMSS 268
      F I+ LA++ ML K Q++E D+ +K GQ FT+L +P T + + V P
Sbjct: 221 PNILLADMSAMLVKAQVSEADVIHLKPGQKAWFTVLGDFLTRYEGQIKDVL+----- 273

Query: 269 GGYNSSTDTASNAVYYIARSFVFNPDGKLATGMTTONTVEIDGVKNVLLIPSLTVKNRGG 328
      + + + + +A++YYAR VFNFG L L MT Q + + + VRNVL IP + + G
Sbjct: 274 -----TEKVNDAIFYARFEVFNFGLLRLDMTAQVHQLTLDVKNVLTIPSLALGDFVG 328

Query: 329 KAFVRV-LGADGKAVEREIRTGMKDSMNTVEVSKLKEGDKVVISSE 372
      +V L +G+ ERE+ G ++ + E+ GL+ GD+VVI E
Sbjct: 329 DNRYKVKLLNGETREREVTIGARNDTDEIVKGLGADGEVVI 373

```

40 Based on this analysis, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF85-1 (40.4kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 19A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein
 45 was used to immunise mice, whose sera were used for Western blot (Figure 19B), FACS analysis (Figure 19C), and ELISA (positive result). These experiments confirm that ORF85-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 92

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 773>:

```

50      1 ..ATTCCGCGCA CGATGACATT TGAACGCGAG GCGAATGCTT ACAAAATCGT
      51 TTGAGACGATT AAGTGCCTCG TATACATAT CCGTTTCGAG TCCGGCGGTA
      101 CGSTTGTGCG CAATACCTCG CACCTTACCT ACTATAGAGA CATAAGCAGC
      151 GGCACAACTGT ATGCGGAGGC CAATTGCGCC GACGCGACGG TAACCTACGG
      201 CAAAGCGGGC GAGAGCAAAA CCGAGCAAG CCCAAGGCT ATGAGTTTGT

```

251 TCACCGTTGC CTGGCAGTTG GCGGCAAAATG ACGCGAAACT CCGCCGGGG
 301 CTGAAAATCA CCAACGGCAA AAAACTTTAT TCCGTCGGCG GTTTGAATAA
 351 GCGGGTACA GGAATAACA GCATAGCGCG CGTGAAACCG GAATCGTCA
 401 AATATCGGCT GCGGCGCGGC GACGATGCGG TAATGTATTT CTTCGACCG
 451 TCCCTGAACA ATATTCGCGC ACAATCGGCG TATACCGAGC ACGGCAAAAC
 501 CTATACGCTG AAATCAAAAT CGGTGCAGAT CAACGGCCAG GCAGCCRAAC
 551 CGTAA

This corresponds to the amino acid sequence <SEQ ID 774; ORF120>:

1 ..IPATMTFERS GNAYKIVSTI KVPLYNIRFE SGGTVVGNLT HPTYYRDIRR
 10 51 GKLYAEAKFA DGSVYTGKAG ESKTEQSPKA MDLFTLAWQL AANDAKLPFG
 101 LKITNGKKLY SVGGLNKAGT GKYSIGGVET EVVKYRVRRG DDAVMYFFAP
 151 SLNNIPAQIG YTTDDGKTYTL KLSVQINGQ AAKP*

Further work revealed the complete nucleotide sequence <SEQ ID 775>:

1 ATGATGAAGA CTTTAAAAA TATATTTTCC GCGGCCATTT TGTCGCGCGC
 15 51 CCTGCGGTGC GGTATGCGG CAGGGCTGCG CCAATCCGCC GTGCTGCATC
 101 ATTCCGGCAG CTACGGCATT CCGGCCAGCA TGACATTTGA ACGCAGCGCG
 151 AATGCTTACA AAATCGTTTC GACGATTAAG GTCCGCTAT ACAATATCCG
 201 TTTCCAGTCC GCGGCTACGG TTGTCCGCAA TACCTCTGAC CTTACCTACT
 251 ATAGAGACAT ACGCAGGGGC AAACCTGTAT CGGAAGCCAA ATTCCGCGAC
 301 GGCAGCGTAA CTTACGGCAA ACGCGGCGAG AGCAAAACCG AGCAAAAGCC
 351 CAAGGCTATG GATTTGTTC CGCTTGCGTG GCAGTTGGCG GCAATGAGC
 401 CGAAATCCCG CCGGGGCGTG AAAATCACCA ACGGCAAAA ACTTTATTCG
 451 GTCGCGGTTT TGAATAAGGC GGGTACAGGA AAATACAGCA TAGGCGCGGT
 501 GGAAACCGAA GTGCTCAAA ATCGGGTGGG GCGCGGCGAC GATGCGGTAA
 25 551 TGTATTTCTT CGCACCGTCC CTGAACAATA TTCCGGCACA AATCGGCTAT
 601 ACCGACGACG GCAAAACCTA CTGACTGAAA CTCAATCCG TGCAGATCAA
 651 CGCGCAGGCA GCCAAACCTG AA

This corresponds to the amino acid sequence <SEQ ID 776; ORF120-1>:

1 MMKTFKNIFS AAILSAALPC AYAAGLPQSA VLHYSYSGYI PATMTFERSG
 30 51 NAYKIVSTIK VELYNIRFES GGTVVGNLT HPTYYRDIRG KLYAEAKFAD
 101 GSVYTGKAGE SKTEQSPKAM DLFTLAWQL AANDAKLPGL KITNGKKLYS
 151 VGGLNKAGTG KYSIGGVETE VVKYRVRRG DDAVMYFFAPS LNNIPAQIGY
 201 TDDGKTYTLK LKSVDINGQA AKP*

Computer analysis of this amino acid sequence gave the following results:

35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF120 shows 92.4% identity over a 184aa overlap with an ORF (ORF120a) from strain A of *N.meningitidis*:

						10	20	30
40	orf120.pep					IPATMTFERSGNAYKIVSTIKVPLYNIRFE		
	orf120a	SAAILSAAALPCAYAGLPKSAVLHYSYSGYIPATXXXXXXXXNAXKIVSTIKVPLYNIRFE						
		10	20	30	40	50	60	
45	orf120.pep	SGGTVVGNLTLPHTYYRDIRRGKLYAEAKFADGSVYTGKAGESKTEQSPKAMDLTFLAWQL						
	orf120a	SGGTVVGNLTLPHTYYRDIRRGKLYAEAKFADGSVYTGKAXXXXXXSPKAMDLTFLAWQL						
		70	80	90	100	110	120	
50	orf120.pep	AANDAKLPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE EVVKYRVRRGDDAVMYFFAP						
	orf120a	AANDAKLPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE EVVKYRVRRGDDAVMYFFAP						
		130	140	150	160	170	180	
55								

-433-

```

              160      170      180
orf120.pep  SLNNIPAIQIGYTDGKTYTLKLSVQINGQAAPKX
              |||
orf120a     SLNNIPAIQIGYTDGKTYTLKLSVQINGQAAPKX
5           190      200      210      220

```

The complete length ORF120a nucleotide sequence <SEQ ID 777> is:

```

1  ATGATGAAGA  CTTTAAAAA  TATATTTC  GCGCCATT  TGTCCGCCG
51  CCTGCGGTGC  GGGATGCGG  CAGGGCTGC  CNAATCCGC  GTGCTGCAT
101 ATTCCGCGCAG  CTACGGCATT  CCGGCCACNA  NNANNTNNGN  ACNNNGNGNC
151 AATGCTTINCA  AAATCGTTTC  GAGGATTAAA  GTGCGCTAT  ACAATATCG
201 TTTGAGTCC  GGGCGTACGG  TTGTGCGCAA  TACCTCGCAC  CCTACCTACT
251 ATAGAGACAT  ACGCAGGGGC  AACTGTATG  CGGAAGCCAA  ATTGCGCGAC
301 GCGACGCTAA  CCTACGGCAA  AGCGGNNNNN  ANCNNNNNNG  NSCAAGAGCC
351 CAAGGCTATG  GATTGTTTCA  CGCTTGCTG  CCAATTGGCG  GCAATATGAG
401 CGPACTCC  CCGCGGCTG  AATACACCA  ACGGCARAA  ACTTATTCTC
451 GTCCGCGTT  TGAATAAGC  GTGTACAGGA  AATACAGCA  TAGGCGGCT
501 GGAAACCGAA  GTGCTCAAT  ATCGGTCGG  GCGCGGCAC  GATGCGGTAA
551 TGTATTCTT  CGCACCGTCC  CTGAACAATA  TTCGCGACAA  AATCGGCTAT
601 ACCGACGAC  GCAAAACCTA  TAGCTGAAA  CTCAAATCG  TGCAGATCAA
651 CGGCCAGGCA  GCCAAACCGT  AA

```

This encodes a protein having amino acid sequence <SEQ ID 778>:

```

1  MKTFKNIFS  AAILSAALPC  AYAAGLPKSA  VLHYSGSYGI  PATXXXXXX
51  NAKKIVSTIK  VPLYNIRFES  GGTVVGNTLH  PTYRDIRRG  KLYAEAKFAD
101 GSVTYGKAXX  XXXXQSPKAM  DLFTLAWQLA  ANDAKLPFGL  KITNGKKLYS
151 VGLLNKAGTG  KYSIGGVETE  VVKYRVRRGD  DAVMYFFAPS  LNNIPAIQIGY
201 TDDGKTYTLK  LKSQINGQA  AKP*

```

ORF120a and ORF120-1 show 93.3% identity in 223 aa overlap:

```

              10      20      30      40      50      60
orf120a.pep  MMKTFKNIFSAAILSAALPCAYAAAGLPKSAVLHYSGSYGIIPATXXXXXXNAKKIVSTIK
              |||
orf120-1     MMKTFKNIFSAAILSAALPCAYAAAGLPQSAVLHYSGSYGIIPATMTFERSGNAYKIVSTIK
              10      20      30      40      50      60
              70      80      90      100     110     120
35  orf120a.pep  VPLYNIRFESGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSGVTYGGKAXXXXXQSPKAM
              |||
orf120-1     VPLYNIRFESGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSGVTYGGKESKTEQSPKAM
              70      80      90      100     110     120
40  orf120a.pep  DLFTLAWQLAANDAKLPPLKITNGKKLYSVGGLNKAGTGKYSIGGVETEYVVKYRVRRGD
              |||
orf120-1     DLFTLAWQLAANDAKLPPLKITNGKKLYSVGGLNKAGTGKYSIGGVETEYVVKYRVRRGD
              130     140     150     160     170     180
45  orf120a.pep  DAVMYFFAPSLNNIPAIQIGYTDGKTYTLKLSVQINGQAAPKX
              |||
orf120-1     DAVMYFFAPSLNNIPAIQIGYTDGKTYTLKLSVQINGQAAPKX
              190     200     210     220
50  orf120a.pep  DAVMYFFAPSLNNIPAIQIGYTDGKTYTLKLSVQINGQAAPKX
              |||
orf120-1     DAVMYFFAPSLNNIPAIQIGYTDGKTYTLKLSVQINGQAAPKX
              190     200     210     220

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF120 shows 97.8% identity over 184 aa overlap with a predicted ORF (ORF120ng) from *N.gonorrhoeae*:

```

55  orf120.pep  IPATMTFERSGNAYKIVSTIKVPLYNIRFE 30
              |||
orf120ng     SAAILSAALPCAYAAARLPQSAVLHYSGSYGIIPATMTFERSGNAYKIVSTIKVPLYNIRFE 69
              |||
orf120.pep  SGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSGVTYGGKESKTEQSPKAMDFTLAWQL 90
              |||
60  orf120ng     SGGTVVGNTLHPAYKDIRRGKLYAEAKFADGSGVTYGGKESKTEQSPKAMDFTLAWQL 129

```

```

orf120.pep  AANDAKLPGLKITNGKKLYSVGGLNKAGTGYKYSIGGVETEYVVKYRVRGGDDAVMYFFAP  150
            |||||||:|||||
orf120ng    AANDAKLPGLKITNGKKLYSVGGLNKAGTGYKYSIGGVETEYVVKYRVRGGDDTVTYFFAP  189
            |||||||:|||||

orf120.pep  SLNNIPAIQIGYTDGKTYTLKLSVQINGQAAKP  184
            |||||||:|||||
orf120ng    SLNNIPAIQIGYTDGKTYTLKLSVQINGQAAKP  223
            |||||||:|||||

```

The complete length ORF120ng nucleotide sequence <SEQ ID 779> is:

```

10      1  ATGATGAAGA  CTTTAAAAA  TATATTTTCC  GCGCCATT  TGTCGCGCCG
51      51 CCGTCCGTCG  GCGTATGCGG  CAAGGCTACC  CCAATCCGCG  GTGCTGCAGC
101     101 ATTCCGGCAG  CTAOGGCATT  CCGCCACAGA  TGACATTGGA  ACGCAGCGGC
151     151 AATGCTTACA  AAATCGTTTC  GAGGATTAAA  GTGCGGCTAT  ACAATATCCG
201     201 TTTGGAATCC  GCGGATACGG  TTGTGCGCAA  TACCTGCGAC  CTGCGTACT
15      251  ATAAAGCAGT  ACGCAGGGGG  AACTGTGTAT  CGGAAACCAA  ATTGCGGCGA
301     301 GCGAGCGTAA  CTTACGGCAA  AGCGGGGCGA  ACGAAACGCC  ACGAAACGCC
351     351 CABGGCTATG  GATTGCTTCA  CGCTTGCCCT  GCAGTTGGCG  GCAAATGAGG
401     401 CGAAACTCCC  CCGGGGCTCG  AAAATCACC  ACGGCAAAA  ACTTTATCCC
451     451 GTCGGGCGCG  TGAATAAGGC  GGGTACGGGA  AAATACAGCA  TagcGGCGGT
20      501  GGAACCGCAA  GTCGTCAAAT  ATCGGGTGGC  GCGCGGGCAG  GATACGGTAA
551     551 CGTATTTCTT  CGCACCGTCC  TTGAACATA  TTCCGGGACA  AATCGGCTAT
601     601 ACGCAGCAGC  GCAAACCTTA  TACGCTGARG  CTCAAATCGG  TGCAGATCAA
651     651 CGGACAGGCG  GCCAAACCGT  AA

```

This encodes a protein having amino acid sequence <SEQ ID 780>:

```

25      1  MNKTFKNIFS  AAILSAALPC  AYAARLPQSA  VLHYSGSYGI  PATMTFERSG
51      51 NAYKIVSTIK  VPLYNIRFES  GGTVVGNTLH  PAYYKDIRRG  KLYAEAKFAD
101     101 GSVTYGKAGE  SKTEQSPKAM  DLFTLAWOLA  ANDAKLPGL  KITNGKKLYS
151     151 VGLLNKAGTG  KYSIGGVETE  VVKYRVRGD  DTVTYFFAPS  LNNIPAIQIG
201     201 TDDGKTYTLK  LKSVQINGQA  AKP*

```

30 In comparison with ORF120-1, ORF120ng shows 97.8% identity in 223 aa overlap:

```

                                10      20      30      40      50      60
orf120-1.pep  MNKTFKNIFS AAILSAALPC AYAAGLPQSA VLHYSGSYGI PATMTFERSG NAYKIVSTIK
            |||||||:|||||:|||||:|||||:|||||:|||||
35      orf120ng  MNKTFKNIFS AAILSAALPC AYAARLPQSA VLHYSGSYGI PATMTFERSG NAYKIVSTIK
            10      20      30      40      50      60

                                70      80      90      100     110     120
orf120-1.pep  VPLYNIRFES GGTVVGNTLH PTYYRDIRRG KLYAEAKFAD GSVTYGKAGE SKTEQSPKAM
            |||||||:|||||:|||||:|||||:|||||:|||||
40      orf120ng  VPLYNIRFES GGTVVGNTLH PTYYRDIRRG KLYAEAKFAD GSVTYGKAGE SKTEQSPKAM
            70      80      90      100     110     120

                                130     140     150     160     170     180
orf120-1.pep  DLFTLAWOLA ANDAKLPGL KITNGKKLYSVGGLNKAGT GRYKYSIGGVETE VVKYRVRGD
            |||||||:|||||:|||||:|||||:|||||:|||||
45      orf120ng  DLFTLAWOLA ANDAKLPGL KITNGKKLYSVGGLNKAGT GRYKYSIGGVETE VVKYRVRGD
            130     140     150     160     170     180

                                190     200     210     220
orf120-1.pep  DAVMYEFASLNNIPAIQIGYTDGKTYTLKLSVQINGQAAKP
            |||||||:|||||:|||||:|||||
50      orf120ng  DTVTYFFAPSLNNIPAIQIGYTDGKTYTLKLSVQINGQAAKP
            190     200     210     220

```

This analysis, including the presence of a putative leader sequence in the gonococcal protein suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 93

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 781>:

```

1 ATGTATCGGA GGAAGGGCG GGCATCAAG CCGTGGATGG GTGCCGGTGC
51 .GGCGTTTGGC GCCTTGGTCT GCGCTGGT TTTT CGCGCTCGGC GATACTTTGA
101 CTCCGTTTGC GGTTCGGCGC GTGCTGGCGT ATGATTATGGA CCCTTTGGTC
151 GAATGGTTGC AGAAAAAGGG TTGAACCGT GCATCCGCTT CGATGCTCTGT
5 GATGGTGTTT TCCTTGATT TGTGTGTGGC ATTATTGTG ATTATCGTCC
251 CTATGCTGGT CGGGCAGTTC AACAAATTGG CATCGCGCTT GCCCAATTA
301 ATCGGTTTGA TGCAGAACAC GCTGCTGCCG TGGTTGAAAA ATACAATCGG
351 CGGATATGTG GAAATCGATC AGGCATCTAT TATTGCGTGG CTTCAGGCGC
401 ATACGGGAGA GTTGAGCAAC GCGCTTAAGG CGTGGTTTCC CGTTTGTATG
10 AGGCAGGGCG GCAATATT..

```

This corresponds to the amino acid sequence <SEQ ID 782; ORF121>:

```

1 MYRRKGRGIK PWMGAGXAFALVVLVFAIGD TLTPFAVAAYLAVLDPLV
51 EWLQKKGGLNR ASASMSVMVF SLILLALLL IIVPMLVGF NNLASRLPQL
101 IGFQNTLLP WLKNTIGGYV EIDQASIIAW LQAHTEGELS ALKAWFPVLM
151 RQGGNI..

```

Further work revealed the complete nucleotide sequence <SEQ ID 783>:

```

1 ATGTATCGGA GGAAGGGCG GGCATCAAG CCGTGGATGG GTGCCGGTGC
51 .GGCGTTTGGC GCCTTGGTCT GCGCTGGT TTTT CGCGCTCGGC GATACTTTGA
101 CTCCGTTTGC GGTTCGGCGC GTGCTGGCGT ATGATTATGGA CCCTTTGGTC
151 GAATGGTTGC AGAAAAAGGG TTGAACCGT GCATCCGCTT CGATGCTCTGT
201 GATGGTGTTT TCCTTGATT TGTGTGTGGC ATTATTGTG ATTATCGTCC
251 CTATGCTGGT CGGGCAGTTC AACAAATTGG CATCGCGCTT GCCCAATTA
301 ATCGGTTTGA TGCAGAACAC GCTGCTGCCG TGGTTGAAAA ATACAATCGG
351 CGGATATGTG GAAATCGATC AGGCATCTAT TATTGCGTGG CTTCAGGCGC
401 ATACGGGAGA GTTGAGCAAC GCGCTTAAGG CGTGGTTTCC CGTTTGTATG
451 AGGCAGGGCG GCAATATTGT CAGCAGTATC GGCACCTGTC TGCTGCTTCC
501 CTTGCTGCTT TACTATTGCC TGCTGGATTG GCAGCGGTGG TCGTGCGGCA
551 TTGCAAACT GGTTCGGAAG CGTTTGGCGG GTGCTTATAC GCGCATACCA
601 GGCAATTGGA ACGAGGTATT GGGGGAATTT TTGGCGGGG AGCTTCTGGT
301 AATGCTGATT ATGGGCTTGG TTACGGTTT GGGATTGGT CTGGCTGGGG
701 TGATTCGGGG GTTGGCATC GGTATGCTTG CCGGATTTT GGTGTTTGTC
751 CTTATCTGG GGTGCTTAC GGGATGTCG CTTGCGACG TCGCGCTT
801 GCTCCAGTTC GGTTCGTGGA ACGGCATCCT ATCGGTTTGG GCGGTTTGTG
851 CCGTAGGACA GTTCTCGAA AGTTTPTTCA TTACGCCGAA AACTGTTGGA
901 GACCGTATCG GCGTGTGCCG GTTTTGGGTT ATCTTTCCG TGATGGCGCT
951 CGGCGAGCTG ATGGGCTTTG TCGGAATGTT GCGGGGATTG CTTTGGCGC
1001 CCGTAACCTT GGTCTTGCTT GCGAGGGCGG TGCAGAAATA TTTTCCGCGC
1051 AGTTTTTACC GGGCGAGGTA G

```

This corresponds to the amino acid sequence <SEQ ID 784; ORF121-1>:

```

40 1 MYRRKGRGIK PWMGAGAAFAALVVLVFAIGD TLTPFAVAAYLAVLDPLV
51 EWLQKKGGLNR ASASMSVMVF SLILLALLL IIVPMLVGF NNLASRLPQL
101 IGFQNTLLP WLKNTIGGYV EIDQASIIAW LQAHTEGELS ALKAWFPVLM
151 RQGGNIVSSI GNLLLLLELL YYFLLDWQRW SCGIKLVPR RFAGAYTRIT
451 GNLEVLGEF LRGOLLVMLI MGLVYGLGLV LVGLDSGFAI GMLAGILVFE
251 PYLGAFTGLL LATVAALLQF GSWNGILSVW AVFAVGQFLE SFFITPKIVG
301 DRIGLSPFWV IFSLMAFGQL MGFVGLAGL PLAAVTLVLL REGVQKYFAG
351 SFYRGR*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

50 ORF121 shows 98.7% identity over a 156aa overlap with an ORF (ORF121a) from strain A of *N. meningitidis*:

```

10 20 30 40 50 60
orf121.pep MYRRKGRGIK PWMGAGXAFALVVLVFAIGD TLTPFAVAAYLAVLDPLV EWLQKKGGLNR
55 orf121a MYRRKGRGIK PWMGAGAAFAALVVLVFAIGD TLTPFAVAAYLAVLDPLV EWLQKKGGLNR
10 20 30 40 50 60
orf121.pep ASASMSVMVFS LILLALLL IIVPMLVGF NNLASRLPQL IGFQNTLLP WLKNTIGGYV
70 80 90 100 110 120

```

-436-

5	orf121a	ASASMSVMVFSLILLALLLIIVPMLVGVGFNNLASRLPOLIGFMQNTLLPWLKNTIGGVV	70	80	90	100	110	120
10	orf121.pep	EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNI	130	140	150			
15	orf121a	EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW	130	140	150	160	170	180
20	orf121a	SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLDSGFAT	190	200	210	220	230	240

The complete length ORF121a nucleotide sequence <SEQ ID 785> is:

15	1	ATGTATCGGA	GGAAAGGGCG	GGGCATCAAG	CCGTGGATGG	ATGCGCGTGC
	51	GGCGTTTGCC	GCCTTGGTCT	GGCTGGTTTT	CGGCTCGGCG	GATACTTTGA
20	101	CTCCGTTTGC	GCTTGGCGGG	GTCGTGGCGT	ATGATTTGGA	CCCTTTGGT
	151	GAATGGTTGC	AGAAAAGGG	TTTGAACCGT	GCATCCGCTT	CGATGCTGT
25	201	GATGGTGT	TCCTTGATT	TGTTGTGGC	ATTATTGTG	ATTATTGTCC
	251	CTATGCTGGT	CGGGCAGTTC	AACAATTTGG	CATCGCGCT	GCCTCAATTA
30	301	ATCGGTTTTA	TGCAGAACAC	GCTGCTGCCG	TGGTTGAAAA	ATACAATCGG
	351	CGGATATGTG	GAATCGATC	AGGCATCTAT	TATTGCGTGG	CTTCAGGCGC
35	401	ATACGGGCGA	GTTGAGCAAC	CGCCTTAAGG	CGTGGTTTCC	CGTTTGTATG
	451	AGGCAGGCGC	GCAATATTGT	CAGCAGTATC	GGCAACTGCG	TGCTGCTTCC
40	501	CTTGCTGCTT	TACTATTTC	TGCTGGATTG	GCAGCGGTGG	TGCTGCGGCA
	551	TTGCGAACT	GTTTCGAGG	CGTTTTCGCG	GTGCTTATAC	GCGCATTACA
45	601	GGCAATTTGA	ACGAGGTATT	GGGCGAATTT	TTGCGCGGCG	AGCTTCTGTT
	651	GATGCTGATT	ATGGGTTTGG	TTTACGGCTT	GGGTTTGGTG	CTGGTCGGGC
50	701	TGGAATCGGG	GTTTGCAATC	GGTATGGTTG	CCGGTATTTT	GTTTTTTTGT
	751	CCCTATTGGG	CGCGGCTTAC	AGGACTGCTG	CTGGCAACCG	TCGCGCGCTT
55	801	GCTCCAGTTC	GTTTCGTGGA	ACGGCATCTT	GGCTGTTTGG	CGCGTTTTTG
	851	CGTAGAGACA	GTTTCGAGG	AGTTTTTCA	TTACGCCGAA	AATCGGCGGA
60	901	GCACGTATGC	GCTTTCGCG	GTTTTCGCT	ATCTTTCGCT	TGATGGCGTT
	951	CGGCGAGCTG	ATGGGCTTTG	TCGGAATGTT	GGCCGGATGG	CCTTTGCGCG
65	1001	CCGTAACCTT	GCTCTTGCTT	CGCGAGGCGG	TGCAGAAATA	TTTTCGCGCG
	1051	AGTTTTTACC	GGGCGAGGTA	G		

This encodes a protein having amino acid sequence <SEQ ID 786>:

40	1	MYRRKGRGIK	PWMDAGAAFA	ALVWLVFALG	DTLTFFAFAA	VLAYVLDPLV
	51	EWLQKKGLNR	ASASMSVMVF	SLILLALLLI	IIVPMLVGVF	NNLASRLPOL
45	101	IGFMQNTLLP	WLKNTIGGVV	EIDQASIIAW	LQAHTGELS	NALKANFPVLM
	151	RQGGNIVSSI	GNLLLLPLLL	YYFLLDWQRW	SCGIAKLVPR	RFAGAYTRIT
50	201	GNLNEVLGEF	LRGQLLVMLI	MGLVYGLGLV	LVGLDSGFAT	GMVAGILVEV
	251	PYLGAFTGLL	LATVAALLQF	GSWNGILAVN	AVFAVGQFLE	SFFITPKIVG
55	301	DRIGLSPFWV	IFSLMAFGQL	MGFVQMLAGL	PLAAVTLVLL	REGVKQYFAG
	351	SFYRGR*				

45 ORF121a and ORF121-1 show 99.2% identity in 356 aa overlap:

50	orf121a.pep	10	20	30	40	50	60
		MYRRKGRGIK	PWMDAGAAFA	ALVWLVFALG	DTLTFFAFAA	VLAYVLDPLV	EWLQKKGLNR
55	orf121-1	10	20	30	40	50	60
		MYRRKGRGIK	PWMDAGAAFA	ALVWLVFALG	DTLTFFAFAA	VLAYVLDPLV	EWLQKKGLNR
60	orf121a.pep	70	80	90	100	110	120
		ASASMSVMVFS	LILLALLLI	IIVPMLVGVF	NNLASRLPOL	IGFMQNTLLP	WLKNTIGGVV
65	orf121-1	70	80	90	100	110	120
		ASASMSVMVFS	LILLALLLI	IIVPMLVGVF	NNLASRLPOL	IGFMQNTLLP	WLKNTIGGVV
70	orf121a.pep	130	140	150	160	170	180
		EIDQASIIAWL	QAHTGELS	NAKAWFPVLM	RQGGNIVSS	IGNLLLLPL	LLYYFLLDWQRW
75	orf121-1	130	140	150	160	170	180
		EIDQASIIAWL	QAHTGELS	NAKAWFPVLM	RQGGNIVSS	IGNLLLLPL	LLYYFLLDWQRW
80	orf121a.pep	190	200	210	220	230	240
		SCGIAKLVPR	RFAGAYTRIT	GNLNEVLGE	FLRGQLLV	MLIMGLVY	GLGLVLVGLDSGFAT
85	orf121-1	190	200	210	220	230	240
		SCGIAKLVPR	RFAGAYTRIT	GNLNEVLGE	FLRGQLLV	MLIMGLVY	GLGLVLVGLDSGFAT

-437-

	orf121-1	SCGI AKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVVLGLDSGFAI	
		190 200 210 220 230 240	
5	orf121a.pep	GMVAGILVFPVYLGAFITGLLATVAALLQFGSWNGILAVWAVFAGVQFLESFFITPKIVG	
	orf121-1	GMLAGILVFPVYLGAFITGLLATVAALLQFGSWNGILSVWAVFAGVQFLESFFITPKIVG	
		250 260 270 280 290 300	
10	orf121a.pep	DRIGLSPFWVIFSLMAFGQLMGFVQMLAGLPLAAVTLLVLLREGVQKRYFAGSFYRGX	
	orf121-1	DRIGLSPFWVIFSLMAFGQLMGFVQMLAGLPLAAVTLLVLLREGVQKRYFAGSFYRGX	
		310 320 330 340 350	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF121 shows 97.4% identity over a 156 aa overlap with a predicted ORF (ORF121ng) from

N.gonorrhoeae:

20	orf121.pep	MYRRKGRGIKPMWGAGXAFALVWLVLGDTLTFFAVAVALVLDPLVEWLQKKGLNR	60
	orf121ng	MYRRKGRGIKPMWGAGXAFALVWLVLGDTLTFFAVAVALVLDPLVEWLQKKGLNR	60
25	orf121.pep	ASASMSVMVFSLLILLALLIIVPMLVGQFNLASRLPOLIGFMQNTLLPWLKNTIGGV	120
	orf121ng	ASASMSVMVFSLLILLALLIIVPMLVGQFNLASRLPOLIGFMQNTLLPWLKNTIGGV	120
	orf121.pep	EIDQASIIAWLQAHTELSNALKAWFPVLMKQGGNI	156
30	orf121ng	EIDQASIIAWFQAHTELSNALKAWFPVLMKQGGNIVSTIGNLLPLLLLYFLLDWHWR	180

An ORF121ng nucleotide sequence <SEQ ID 787> was predicted to encode a protein having amino acid sequence <SEQ ID 788>:

	1	MYRRKGRGIK FWMGAGAAFA ALVWLVLGDTLTFFAVAVALVLDPLV	
	51	EWLQKKGLNR ASASMSVMVF SLILLALLI IIVPMLVGQFN LASRLPOL	
35	101	IGFMQNTLLP WLKNTIGGV EIDQASIIAW FQAHTGELS ALKAWFPVLM	
	151	KQGGNIVSTI GNLLPPLLL YFFLLDWHWR SCGIPKLVPR RFAGAYTRIT	
	201	GNIAKIVWGKF LRGQLLGETE RGAUVCRVGR ECWEGGGARS RPSDDGWPRW	
	251	GGG*	

Further work revealed the following gonococcal DNA sequence <SEQ ID 789>:

40	1	ATGTATCGGA GAAAGGACG GGGCATCAAG CGTGGATGTG GTGCCGCGCG	
	51	GGCGTTTGCC GCGTTGGTCT GGCTGGTTTA CGCGCTCGGC GATACCTTGA	
	101	CTCCGTTTGC GGTTCGGCGC GTGCTGGGCT ATGTGTGTGA CCTTTTGGTC	
	151	GAATGGTTGC AGAAAAAGG TTTGAACCGT GCATCCGCTT CGATGCTCTGT	
	201	GATGGGTGTT TCCTGATTT TGTGTTGGC ATTPTGTTG ATTAATGTCC	
45	251	CTAPCTGCT CGGCAGTTC AATAATTTG CATCTCGCCT GCCCAATTA	
	301	ATCGGTTTGA TGCAGACAC GCTGCTGCGC TGGTTGAAAA ATACATCCG	
	351	CGGATATGFG GAAATCGATC AGGCATCTAT TATTGCGTGG TTTACGCGCG	
	401	ATACGGGGGA GTTGAGCAAC CGCCTTAAGC COTGTTTCC CGTTTGTATG	
	451	AAACAGGGCG CCAATATTGT CAGCATATC GGCACCTCGT TGCTGCGCGC	
50	501	CTTCTGCTCT TACTATTTC TGCTGGATTG GCAGCGGGTG TCGTGCGGCA	
	551	TGCGCAAACT GGTTCGAGG CGTTTTCGCC GTGCTTATAC GCGCATTAGC	
	601	GGTAATTGTA ACGAGGTATT GGGCGAATT TTTGCGGGTC AGCTTCTGGT	
	651	GATGCTGATG ATGGGCTTGG TTTACGGTTT GGGATTGATG CTAGTCGAGC	
	701	TGGATTGCGG ATTTGCCATC GGTATGGTTG CGGATATTTT GGTGTTTGTC	
55	751	CCCTATTGTT GTGGCTTTAC GGGATTGCTG CTTCGCACTG TTGACGCCCT	
	801	CTCCAGTTC GGTTCGTGGA ACGGAATCTT GCGTGTTTGG GCGGTTTTTG	
	851	CGTCGCTCA GTTCTCGGAA AGTTTTTCA TTACGCGGAA AATTGTAGGA	
	901	GACCGTATCG GCGTGTGCG GTTTTGGGTT ATCTTTTTCG TGATGGCGTT	
	951	CGGAGAGCTG ATGGGCTTTG TCGGAATGTT GGCAGGATTT CTTTGGCGCG	
60	1001	CGTAACCTCT GGTCTTGTCT CGGAGGGCG CGCAGAAATA TTTTGGCGCG	
	1051	AGTTTTTACC GGGCGAGGTA G	

This corresponds to the amino acid sequence <SEQ ID 790; ORF121ng-1>:

	1	MYRRKGRGRIK	PWMGAGAAFA	ALVWLIVYALG	DTITPFFAVAA	VLAYLDPLV	
	51	EWLQKQGLNR	ASASMSVMVF	SLILLALLLL	IIVPMLVGQF	NNLASRLPQL	
	101	IGFMONTLLP	WLKNTIGGYV	EIDQASIIAW	FOAHTGELSN	ALKAWFPVLM	
5	151	KQGNIVSSI	GNLLLEPLL	YFLLDWQRW	SCGIAKLVPR	RFAGAYTRIT	
	201	GNLNVLGEF	LRGQLVML	MLVYGLGIM	LVSLDSGFAI	GMVAGTIVF	
	251	PYLGAFTGLL	LATVAALLQ	FGSNGILAVN	AVFAVQGLE	SFFITPKIVG	
	301	DRIGLSPFWV	IFSLMAFGEL	MGFVGMLAGL	PLAAVTLVLL	REGAQKYFAG	
	351	SFYRGR*					
10	ORF121ng-1 and ORF121-1 show 97.5% identity in 356 aa overlap:						
		10	20	30	40	50	60
	orf121-1.pep	MYRRKGRGRIK	PWMGAGAAFA	ALVWLIVYALG	DTITPFFAVAA	VLAYLDPLV	EWLQKQGLNR
	orf121ng-1	MYRRKGRGRIK	PWMGAGAAFA	ALVWLIVYALG	DTITPFFAVAA	VLAYLDPLV	EWLQKQGLNR
15		10	20	30	40	50	60
	orf121-1.pep	70	80	90	100	110	120
	orf121ng-1	ASASMSVMVF	SLILLALLLL	IIVPMLVGQF	NNLASRLPQL	IGFMONTLLP	WLKNTIGGYV
20		70	80	90	100	110	120
	orf121-1.pep	130	140	150	160	170	180
	orf121ng-1	EIDQASIIAW	FOAHTGELSN	ALKAWFPVLM	RQGGNIVSS	IGNLLLP	LYFYFLDWQRW
25		130	140	150	160	170	180
	orf121-1.pep	190	200	210	220	230	240
	orf121ng-1	SCGIAKLVPR	RFAGAYTRIT	GNLNEVLG	EFLRGQLLV	MLVYGLGIM	LVGLDSDGFAI
30		190	200	210	220	230	240
	orf121-1.pep	250	260	270	280	290	300
	orf121ng-1	GMLAGILVF	PYLGAFTGLL	LATVAALLQ	FGSNGILAVN	AVFAVQGLE	SFFITPKIVG
35		250	260	270	280	290	300
	orf121-1.pep	310	320	330	340	350	
	orf121ng-1	DRIGLSPFWV	IFSLMAFGEL	MGFVGMLAGL	PLAAVTLVLL	REGAQKYFAG	SFYRGRX
40		310	320	330	340	350	
	orf121-1.pep	370	380	390	400	410	420
	orf121ng-1	MLWNTISLL	SDLPAMF	-----	NKSNWLLN	LPNYP	ELIDYSMDVS
45		370	380	390	400	410	420
	orf121-1.pep	430	440	450	460	470	480
	orf121ng-1	AWFPVLMK	QGGNIVSS	IGNLLLP	LYFYFLDWQRW		

In addition, ORF121ng-1 shows homology to a permease from *H. influenzae*:

	sp P43969 PERM_HAEIN PUTATIVE PERMEASE PERM_HOMOLOG Length = 349				
	Score = 69.9 bits (168), Expect = 2e-11				
	Identities = 67/317 (21%), Positives = 120/317 (37%), Gaps = 7/317 (2%)				
50	Query: 26	VYALGDTITPFFAVAAVLAYLDPLVLEWL-QKQGLNRASASMSVMVFXXXXXXXXXXV	84		
	Sbjct: 32	IYFFGDILAPILLIALVLSYLLIPINFLNQYLKCPRLATILIFGSGFIAGLAAVFFVLVLP	91		
55	Query: 85	MLVGQFNNLASRLPQLIGFMONTLLPWLKNTIGGYVE-IDQASIIAWFOAHTGELSN	143		
	Sbjct: 92	MLWNTISLLSDLPAMF-----NKSNEWLLNLPNYP	ELIDYSMDVS	147	
	Query: 144	AWFPVLMKQGGNIVSSIGNXXXXXXXXXXXXXQWRWSCGYAKLVPRRFAFGAYTRITGNL	203		
60		+	+	+	N+VS
	Sbjct: 148	SAVKLSLASIMNLVSLGIFYAFVPLMMFM	LKQKSELLQGVSRFLPKNRNLA	FXRWK-EM	206
	Query: 204	NEVLGELRGQXXXXXXXXXXXXXXXXXXXXSGFAI	GMVAGILVF	VEYXXXXXXXXXX	263
		+	+	+	G+V VFY
65	Sbjct: 207	QQQISNYHKGKLEILIVTLITYIIFLIFGLNYP	LLAFVGLSVLVYIGAVITPVA	266	

Query: 264 XXXXXQFGSWNGILAVWAVFVGQFLESFFITPKIVGDRIGLSPFWIFSLMAFGLMGF 323
 QFG + FAV Q L+ + P + + L P + I S++ FG L GF
 Sbjct: 267 LVALFQFGISPTFWYIIIAFAVSQLDGNLLVPLFSEAVNLHPLIIIIISVLIFGGLMGF 326

5 Query: 324 VGMLAGLPLAAVTLVLL 340
 G+ +PLA + ++
 Sbjct: 327 WGVFFAIPLATLVKAVI 343

Based on this analysis, including the presence of a putative leader sequence and transmembrane domains in the two proteins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 94

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 791>:

```

1  ..ACTGCTTTT CGGGCGGCGT GCGCTTGAGT CCATCATGAC TCGTCATATT
51  TTGTGCTCTT GGGGAAACCGT ATCAACAAAC AGCGGCCATC TTAACATTTT
101  TTGTCAGCTC CTGCGCGCGC CGTTCAAATG CGTACCAGCA ATACCGCCGC
151  CTGCGCCTCT ATGCGCTTCCA TCGCGCCGAG ATAGCGGAGT TTTTCTGTGG
201  TTTTGCTCTT GATGTTGAGC CAGGAAATGT CTATGCCCAA ATCGCGCGCG
251  ATGTTGGCAC GCATTTGGGG AATGTGCGGC GCGAGTGTGG GTTCTGTGTC
301  AATCAGCGTC GTATCGACAT TGACGCGCTG CCAACCTGTC GCTTGAACGC
351  TTGTGATACG CGCAGCAAA AGSACGCGCG TGTCGCTCTC TTTGAACCTC
401  GCGCGCGTGT CGGGGAAATG GCTGCCGATA TCGCCAAAC CTGCCGACC
451  GAGCAGCGCG TCGGTAAACG CGTGACGAGC CGCATCGGCA TCGAGTGTG
501  CGAGCAGCCC TTTTTCAAAT GGGATTTCAA CTCGCCAAG TATCAG..

```

25 This corresponds to the amino acid sequence <SEQ ID 792; ORF122>:

```

1  ..TAFSAALRLS PSKLVIFLSF GKPYQQTAAI LTFCTSCPP RSNAYQYRR
51  LRLYAFHPPE IAEFFVGFAP DVDARNVYQ IGGDVGTHLR NVRECGFLC
101  NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGVGEM AADIAQTCT
151  EQRVGNVQQ RIGIGVSEQ FFKWDFNSAK YQ..

```

30 Further work revealed the complete nucleotide sequence <SEQ ID 793>:

```

1  ATATCGTACT GGGCAAGCAG TCGCGCGGAT TTTTGGAGT TAGATACCGC
51  GCCTTTGATT TTTTGGCGC TCTTACCAA GGCTTCGATG AAAAGTTGTA
101  TGGTCGAGCC GGTACOGATG CGGATATATT CATTTTCGGG TACGAATTTCG
151  ACTGCTTTT CGGCGGCGAT GCGCTTGAGT TGTCTTTGTG TCGTCATATT
351  TTTTGCTCTT GGGGAAACGT ATCAACAAAC AGCGGCCATC TTAACATTTT
201  TTTGCAGCTC CTGCGCGCGC CGTTCAAATG CGTACCAGCA ATACCGCGCG
251  CTGCGCCTCT ATGCGCTTCCA TCGCGCCGAG ATAGCGGAGT TTTTCTGTGG
301  TTTTGCTCTT GATGTTGAGC CAGGAAATGT CTATGCCCAA ATCGGGCGGCG
351  ATGTTGGCAC GCATTTGGGG AATGTGCGGC GCGAGTTTGG GTTCTGTGTC
401  AATCAGCTC GTATCGACAT TGACGCGCTG CCAACCTGTC GCTTGAACGC
451  TTTGATACGC CGCAGCAAA AGGACGCGCG TGTCCGCTATC TTTGAACCTC
501  GCGGCGGTGT CGGGGAAATG GCTGCCGATA TCGCCAAAC CTGCCGACC
551  GAGCAGCGCG TCGGTAAACG CGTGACGAGC CGCATCGGCA TCGAGTGTG
601  CGAGCAGCCC TTTTTCAAAT GGGATTTCAA CTCGCCAAG TATCAGCTTT
651  CTGCGCTCGG TCAGTTGGTG GACATCGTAG CCGTTCGGA TACGAGTGT
701  CGTCATCGTT TGCTTCTCTG A
751

```

This corresponds to the amino acid sequence <SEQ ID 794; ORF122-1>:

```

1  ISYWASSSPD FLEVDTAPLI FLELLPKASM KKLMEVPVPM PIYSFGSTNS
51  TAFSAAMRLS SSCVIFLSF GKPYQQTAAI LTFCTSCPP RSNAYQYRR
101  LRLYAFHPPE IAEFFVGFAP DVDARNVYQ IGGDVGTHLR NVRECGFLC
151  NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGVGEM AADIAQTCT
201  EQRVGNVQQ RIGIGVSEQ FFKWDFNSAK YQLSFGQLV DIVALSDDTV
251  RHRLCS*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF122 shows 94.0% identity over a 182aa overlap with an ORF (ORF122a) from strain A of *N.*

meningitidis:

					10	20	30
5	orf122.pep				TAFSAALRLSPSXILVIFLSFGKPYQQTAAI	: : : : : : : : :	
	orf122a	FLPLLPKAMMKLMVEPVMPMYSFSGTNSTAFSAAMLSSSCVVIFLSFGKPYQQTAAI					80
		30	40	50	60	70	
10	orf122.pep	40	50	60	70	80	90
	orf122a	LTFFTSCPFRSNAYQQYRRLRLRYAHFPPEIAEFFVFVGFAFDVDARNVYAIGGGDVGTHLR					
	orf122a	LTFFTSCPFRSNFYQQYRRLRLRYAIIAEITEFFVGFAFXDVDARNVYAIGGGDVGTHLR					
		90	100	110	120	130	140
15	orf122.pep	100	110	120	130	140	150
	orf122a	NMRREGGFCLNHGRIDIDLPTLLRNALLIRATOKDAVRIFEELCGGVGEMAADIACTCRT					
	orf122a	NMRREFGLCNHGRIDIDLPTLLRNALLIRATOKDAVRIFEELCGGVGEMAADIACTCRT					
		150	160	170	180	190	200
20	orf122.pep	160	170	180			
	orf122a	EQRVGNGVOORIGIVGSQPFFKKWDPNRSARYQ					
	orf122a	EQRVGNGVOORIGIVGSQPFFKKWDPNRSARKYLSAQLVDIALSDTVHRHLCSX					
		210	220	230	240	250	

The complete length ORF122a nucleotide sequence <SEO ID 795> is:

	1	ATATCATGAT	GGGACGACG	TTCACTGGAT	TTTTTGGAAG	TAGATACCGC
30	51	GCCTTCATT	TTTTTGCGCG	TCTTACGAA	GGCTTCGAGA	AAAAAGTTGG
	101	TGGTCGAACG	GGTACCGATG	CGGATGATT	CGTTTTCGCG	TACGAAATCG
	151	ACTCGTCAT	CGCGCGCAT	TGCTTGAGT	TGCTTCGTG	TGCTCATATT
	201	TTTGTCTTT	GGGAAACGCT	ATCAACAACG	AGCGCCATC	TTACATATT
	251	TTNNNAGCT	CTCGCCGCGC	GCTTCAACT	CTTACACGA	ATACGCCGCG
35	301	CTGAGCTCT	ATGCTCTTCA	TGGCGCGAG	ATACCAAGCT	TTTTCGTTGG
	351	TTTGGACCT	GANGTGTGAC	CACGAAATGT	CTATGCCCAA	ATCGCGGGG
	401	ATGTTGGCAC	GCAATTTGCGG	AAATACGCGC	CGAGATTGG	TTTTCTGTGC
	451	AATCACGCTG	GATATGACAT	TGCACGCGCT	CCGAACCTCG	GCTCGAAGCG
	501	TTTGATAGC	CGCAGCAAA	AGGAGCGGCG	TGTGCGCATC	TTTGAACCTT
	551	CGCGCGGTGT	CGGGGAAATG	CTGCGGATA	TGCGCCAAAC	TGCGCGCACC
40	601	GAGCAGCGCG	TGCGTAAAGT	CTGACGACG	CGCATCGGCA	TGCGAGTGTG
	651	CGAGCAGCCC	TTTTTCAACT	GGGATTTCAA	CTCGGCCAAG	TATCAGCTTT
	701	CTGCTCTGCG	TCAGTTGTGT	GACATCTGAT	CCCTGTCCGA	TACGAGTGT
	751	CGTACTCGTT	TGCTTCTCT	A		

This encodes a protein having amino acid sequence <SEO ID 796>:

45	1	ISYWASSSLD	FLEVDTPFLI	FLPLLPKASH	KKLMWEEVFN	PMYSFSGTNS
	51	TAFSASMRLS	SCSUVIFLPS	LFPTQYQOTAA	LTFFTSXCPF	RSNYQYQYRR
	101	LRLYAFIAHE	PTTFVVGFAF	XVDARNVVAQ	IGGDVGTHLR	NMRREFGLFC
	151	NHGRIDIDRL	PTFRLNALIR	TRQKDAVRI	FELCGGVGEM	AADIAQTQCR
	201	EQRVNGVQVQ	RIGIGVSEQP	FFXWDFNSAK	YQLSAFGQLV	DIVALSDTDV
50	251	RHRLCS*				

ORF122a and ORF122-1 show 96.9% identity in 256 aa overlap:

[illegible]

		70	80	90	100	110	120
		130	140	150	160	170	180
5	orf122a.pep	XVDARNVYQIIGDVGTHLRNMRRFEGFLCNHGRIDIDRLPTLRNLALIRRTQKDAAVRI					
	orf122-1	DVDARNVYQIIGDVGTHLRNMRRFEGFLCNHGRIDIDRLPTLRNLALIRRTQKDAAVRI					
		130	140	150	160	170	180
10	orf122a.pep	FELCGGVGEMAAIDIAQTCTEQRVGNVQQRIGIGVSEQPFFKWFNSAKYQLSAFGQLV					
	orf122-1	FELCGGVGEMAAIDIAQTCTEQRVGNVQQRIGIGVSEQPFFKWFNSAKYQLSAFGQLV					
		190	200	210	220	230	240
15	orf122a.pep	DIVALSDTVRHLRCSX					
	orf122-1	DIVALSDTVRHLRCSX					
		250					
20							

Homology with a predicted ORF from *N.gonorrhoeae*

ORF122 shows 89.6% identity over a 182 aa overlap with a predicted ORF (ORF122ng) from *N.gonorrhoeae*:

25	orf122.pep	TAFSAALRLSPXLVIFLSFGKPYQQTAAI	30
	orf122ng	FLPLLKASMKIMVEPVPMFMSFSGTNSAFSAAMRLSSSCVVI FLSPGKPYQQTAAI	80
	orf122.pep	LTFFCTSCPSPSNAYQQYRRLRLYAFHPPEIAEFFVGFADVDARNVYQIGSDVGTHLR	90
30	orf122ng	LTFFCTSWPSPSNPYQQYRRLRLYAFHPPEIAEFFVGFADIDARNIDTQIGSDVGTHLR	140
	orf122.pep	NVRRECGFLCNHGRIDIDRLPTLRNLALIRRTQKDAAVRI FELCGGVGEMAAIDIAQTCT	150
	orf122ng	NVRCEFGFLCNHGRIDIDHLP LTLRLNALIRRTQKDAAVRI FELCGGVGEMAAIDIAQTCT	200
35	orf122.pep	EQRVGNVQQRIGIGVSEQPFFKWFNSAKYQ	182
	orf122ng	EQRVGNVQQRVGRIMPEQPFKWFNSAKYQLSAFGQLVDIVALSDTVDIRHLRCS	256

The complete length ORF122ng nucleotide sequence <SEQ ID 797> is:

40	1	ATGTCGTACC	GGGCAAGCAG	TTGCGCGGAT	TTTTTGAGGG	TTGAAACCGC
	51	GCCTTTGATT	TTTTTACCGC	TTTTGCCCAA	GGCTTCGATG	AAGAAATTGA
	101	tgGTCGAACC	GgtACCGATG	CCGATGTATT	CGTTTTCGGG	TACGAATTGC
	151	ACTGCTTTTT	CGGCGGGGAT	GCGCttgAgt	TCgtcttgcg	TcgTCATATT
45	201	TTTATctctt	gGGAaaccct	atcaAcaAac	agccgcacat	TTAACATTTT
	251	TTTGCACTtc	ctggcgccgc	cgctcaAATc	cgtaacacGa	ataccgcccgc
	301	ctgcgcctCT	AtgcCTTCCA	TCGCGCCGAG	ATAGCCGAGT	TTTTTCCTGG
	351	TTTTGCCTTT	GAAtTtGAGC	CACGAATAT	CGatacCCAa	atcgggcgCG
	401	ATGTTGGCAC	GCATTTGGCG	AATGTGCGGT	GGAGTTTGG	GTTTCTGTGC
	451	AATCACGGTC	GTATCGACAT	TGACCACCTG	CCAACCTGCG	GCCTGAACGC
50	501	TTTGATACGC	CGCACGCAAA	AGSAGCGGGC	TGTCGCGATC	TTTGAATCTT
	551	GCGGCGGTGT	CGGGAATATG	GCTGCGGATG	TGCGCCAAAC	CTGCCGACCC
	601	GAGCAGCGcg	tcggtaaCGG	CGTGCGAGCG	cgcgTcgGCa	TCCGAATGCC
	651	CGAGCAGCCC	TTTTTCAAA	GGGATTTCAA	CTCGCCCAAG	TATCATGTTT
	701	CTGCCCTTCG	TCAATTTGGT	GACATCGTAG	CCCTGTCCGA	TACGGATATT
55	751	CGTCATCGTT	TGTGTTCTGT	A		

This encodes a protein having amino acid sequence <SEQ ID 798>:

	1	MSYRASSSPD	FLEVETAPLI	FLPLLKASM	KKLMVEPVPM	PMYSFSGTNS
	51	TAFSAAMRLS	SSCVVI FLSP	GKPYQQTAAI	LTFFCTSWPP	RSNPYQQYRR
60	101	LRLYAFHPPE	IAEFFVGFAG	DIDARNIDTQ	IGGDVGTHLR	NVRCEFGFLC
	151	NHGRIDIDHLP	FTLRNLALIR	RTQKDAAVRI	FELCGGVGKM	AADVAQTCTRT
	201	EQRVGNVQQR	RVGIRMPPEQ	FFKWFNSAK	YQLSAGQLV	DIVALSDTDI
	251	RHLRCS*				

ORF122ng and ORF122-1 show 92.6% identity in 256 aa overlap:

		10	20	30	40	50	60
5	orf122-1.pep	ISYWASSSPDFLEVD	TAPLIFLPLPKASMK	KLMVEPVPMP	PIYSFSGTNSTAF	SAAMRLS	
	orf122ng	MSYRASSSPDFLEVD	TAPLIFLPLPKASMK	KLMVEPVPMP	PIYSFSGTNSTAF	SAAMRLS	
		10	20	30	40	50	60
10	orf122-1.pep	SSCVVIFLSFGKPYQQT	AAILTFCTSCPPRSNA	QQYRRLRLYA	FHPPIAEFFVGF	PAF	
	orf122ng	SSCVVIFLSFGKPYQQT	AAILTFCTSWPPRSNP	QQYRRLRLYA	FHPPIAEFFVGF	PAF	
		70	80	90	100	110	120
15	orf122-1.pep	DVDARNVYAQIGGDVG	THLRNVRFGLCNHGR	IDIRLPTLRNALIR	RTQKDAAVRI		
	orf122ng	DIDARNIDTQIGGDVG	THLRNVRFGLCNHGR	IDIRLPTLRNALIR	RTQKDAAVRI		
		130	140	150	160	170	180
20	orf122-1.pep	FELCGGVGEMAAADIA	QTCRTQQRVNGVQQ	RAIGIGVSEQPF	FKWDFNSAKYQL	SAFGQLV	
	orf122ng	FELCGGVGEMAAADIA	QTCRTQQRVNGVQQ	RAIGIGVSEQPF	FKWDFNSAKYQL	SAFGQLV	
		190	200	210	220	230	240
25	orf122-1.pep	DIVALSDTDVHRRLCS	X				
	orf122ng	DIVALSDTDIRHRLCS	X				
30		250					

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 95

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 799>:

35	1	..GCCGCGCGGA	GTGCGAACAA	CATTTCGCG	CGTTTTGCGG	AAACACCCGT
	51	CGCTGTGAGC	GTTACCCCTGA	TCGGCAGCGT	ACTTGCCGTC	ATGCTGCCGC
	101	TTACCGAATA	TGAAGAACTTC	CTGCTGCTTA	TCGGCTCGGT	ATTTCGCGCG
	151	ATGCGGCGGA	TTTTGATGCG	CGACTTTTTC	GTCTGGAAC	GCGGTGA

This corresponds to the amino acid sequence <SEQ ID 800, ORF125>:

40	1	..AGASANNISA	RFAETPVAVS	VTLIGTVLAV	MLPVEYENF	LLIGSVFAP
	51	MGGPDCRLFR	LETA*			

Further work revealed the complete nucleotide sequence <SEQ ID 801>:

	1	ATGTCGGGCA	ATGCCCTCTC	TCCTTCATCT	TCCTCCGCCA	TCGGGCTGAT
	51	TTGGTTTCGG	GCGCGGATG	CAGATTGCCGA	AATCAGCAGC	GGTACGCTGC
45	101	TTGGCGCTTT	GGGCTGGCAG	CGCGGCTCTG	CGGCTCTACT	TTTGGGTCAT
	151	GCCGCTCGGC	GCGCGCTGTT	TTTTCGGGCG	GCSTATATCG	GCGCACTGAC
	201	CGGACGCAGC	TCGATGGAAA	GCGTGCCTCT	GTCGTCGCG	AAACGCGGTT
	251	CAGTGCTGTT	TTCCGTGGCG	AATATGCTGC	AACCTGGCCG	CTGGAACGGG
	301	GTGATGATTT	ACGCGGCGCG	AACGCTCAGC	TCGCGTITGG	GCAAAAGTGT
50	351	GTGGGACGGC	GAATCTTTTG	TCGTGGGCG	ATTGGCAAC	GCGCGCTGAG
	401	TGTGCTGTG	GCTGGTTTTC	GCGCAGCAGC	AAACAGCGCG	GCTGAAAACC
	451	TTTCGATGC	TGCTGATGCT	GTGGCGGTT	CTGTGCTGTA	GTGCGGAAGT
	501	CTTTTCCAGC	CGAGGACGCA	CCGCGGCACA	GGTTTTCAGC	GCGCATGAST
	551	TCGGAACGCG	AGTCGAGCTG	TCGCGCGTGA	TGCGCGCTTC	CTGGCTGCGG
55	601	CTTGCCGCGC	ACTACACCGG	CCAGCGCGCG	CGCGGCTTTC	CGGCAACCTC
	651	GACGGCAGCG	CTCCCTACGA	CGCTGACCGG	CTCGCTGTGG	TATGCTCTTG
	701	GTTTGGCAGC	GCGGTTGTTC	ATCGAGAA	CCGACGTGCG	AAAATCCTTG
	751	CTGGGCGCAG	GTTTGGGTGC	GCGAGGCATT	TTGGCGGTGC	TCCTCTCCAC

801	CGCTACCAACA	ACGTTTCTCG	ATGCTCATTC	CGCGGGCGG	AGTGGAACA
851	ACATTTCGCG	GGTTTTCGCG	GAAACACGCG	TGCGTCTCGG	CGTTACCTGG
901	ATCGCGACGG	TACTTGCCTG	CA'TGCTGCCC	TAT'ACCGAAT	ATGAAAACTTT
951	CTCGTCTGTT	ATCGGCTGCT	TATTTGGCGG	GAT'GGCGCG	GTTTGTATGG
1001	CGCACTTTT	CGTCTGAACT	GGCGGCTGAG	AGATT'GAAG	CTTTGACCTTT
1051	GGCGGACTGG	TTCTTGAGCT	TGGGGGCTTG	ATCTCTCAAC	CTCTCTCGTG
1101	CTCGAAGCTG	CGTCAAGCTG	CGATACGCTG	CGAAGACGCT	CGAAGACGCT
1151	CGCTGCGCAT	TGCCACCGTA	TGCGTAGCGC	TTTTCTTTAA	AAAAACCCAA
1201	TCCTTATGAA	GGAGACGCTC	ATGAG		

10 This corresponds to the amino acid sequence <SEO ID 802: ORF125-1>:

	1	MSGNASPPSS	SSAIGLWFG	AMSVLEAEST	GTLLAPLPGW	RGLAALLLGH
	51	AGUGALFFAA	YYAIGALTGR	SMESVLSFVA	KRGSLVLSVA	NMLQAGWTA
	101	VMIYATAGTS	SALGVKLWFG	SFVWVALAN	GILALVIMLV	GARTK7GLKT
15		VSMILLMLAV	LWLSAEVEST	AGSTPAWEL	GMSPTAVEL	SAMPTLSWLP
	201	LADYTRHAR	RPPAATLLT	LAYTLTGCMW	YALGSLAARL	TGTDVAKILL
	251	LGLAGLNGAL	LVYLSTVTT	FLTDAYSAGA	SANNISIAAR	TEPVAGVTLT
	301	IGTVMLMLF	LYLVWAPLL	IGSVFAPAF	VLAADIAIV	RREELGFKPT
	351	ASLVNLVGL	ILYVFLSSG	NESSIGLTP	WMSAVAIATV	SVLFEFKFOT
	401	SLRNS*				

20 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF125 shows 76.5% identity over a 51aa overlap with an ORF (ORF125a) from strain A of *N. meningitidis*:

```

25      orf125.pep      AGASANNISARFAETPVAVSVTLIGTVLAV
                        ||:|||||:::| |:|:::|:|
orf125a      KILLGAGLGAAGILAVVLSTVTTFLDAYSAAGVSANNISAKLSEIPIAVAVVGTLLAV
                250      260      270      280      290      300

30      orf125.pep      MLPVTEYENFLLLIGSVFAPMGGDCFRLETX
                        :||| ||||| ||||| :|
orf125a      LLPVTEYENFLLLIGSVFAPMAAULIADFFVLRREEIEG
                310      320      330      340

```

35 The ORF125a partial nucleotide sequence <SEO ID 803> is:

	1	ATGTGCGGGCA	GGCGCTCTCT	TCNHTCATCT	TCGCGCGGCCA	TCGGGCTGAT
	51	TGTGCTGTGAT	AGCGGCTGAT	GTATCCAGCA	ATATCAGCATC	GGATCACTGC
	101	TGGCGCTCTT	GGGCTGGGAC	CGGGCTCTGG	CNCGTCTGTCT	TTTGGGTCAT
40	151	CGCGCTGGCG	GCGCGCTGCG	TTTGGGCGCG	GGGTATATCG	GCGCACTGAC
	201	CGACGACNAC	TCGATGGAAA	GGCGCGCTCT	TCGCTTCCGC	AAACGCGGTT
	251	CAGTGTCTGT	TTGCGTGTGG	AATATGTGCG	AACGTGGCGG	CTGGACGGCG
	301	GTGATGATT	ACGCGGGGCG	AACGCTGATC	TCGCGTCTGG	GCAAACTGTT
	351	GTGGGACGGC	GAATCTTTTG	TGTGTGGGAC	ATTGCAAAAC	GCGCGGTGTA
45	401	TGTGCTGTGT	CGTGTTTGT	GGGCGACAGCA	AAACAGGGCG	GCTGAACACC
	451	GTTTGCATGC	GTGTGATGCT	GTGCGCGTGT	GTGCTGGCTGA	GTCGCGAANT
	501	NTTTTCCAGC	CAGGACGAGCA	CGCGCGCGNT	GGTNNCGAGT	GGCATGAGTT
	551	TGGACAGCGC	AGTCGAGCTG	TCGCGCGTNA	TTCGCCCTTC	TTGCGTCGCG
	601	CTGCGCGCGC	ACTACAGCGG	CCACGCGCGG	CGCCGCTTTG	CGGACACCGT
50	651	GACGCAACGC	CTGCGCTACA	CGCTGACGCG	CTGCTGAGAT	TATGCTCTTC
	701	GTTTGCGGAC	GCGCTGTGTC	ACCGGAGAAA	CCGAGCTGGC	MAAATCTCTG
	751	CTGCGGACGC	CTGCGGCTGC	GCGCAGCAAT	CGCGGCTGCG	TCCTCTCGAC
	801	CGTATACCAC	ACTTTTCTGT	CGCGCGGCTA	CGCGGCGGTA	AGTCCGCGTA
	851	ATATTTCGCG	CAAACTCTGC	GAATATACCA	TCGCGCTTGC	CTGTGCGGTT
55	901	CTGGGCAACG	TGCTTGGCGT	CTCTCGCCGC	PTGACCGAAT	ATGAAACAT
	951	CTCTGCTGCT	ATTGCGCTGG	TATTTGGCGG	AGTGGCGGCG	GTTTGTATTG
	1001	CGCGACTTTT	CSTTCTGAAA	GCGGCTGAGG	AGATTGAAAG	C..

This encodes a protein having the partial amino acid sequence <SEQ ID 804>:

1 MSGNASSXSS SAAIGLIWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH
51 AVGGALFFAA AYIGALTGXX SMESVRLSEG KRGSVLFSSA NMLOLAGWTA

-444-

101 VMIYAGATVS SALGKVLWDG ESEFVWALAN GALIVLWLVF GARRTGGKLT
 151 VSMILMLLV LMLSSEVFT AGSTAAVVD GMSFGTAVEL SAVMPLSWLP
 201 LAADYTRHAR RPPAATITAT LAYTLTGCCWM YALGLAALF TGETDVAKIL
 251 LGAGLGAAGT LAVVLSTVTT TFLDAYSAGV SANNISAKLS EPIAVAVAV
 5 301 VGTLLAVLLP VTEYENFLLL IGSVFAPMAA VLIADFFVLK RREEIEG..

ORF125a and ORF125-1 show 94.5% identity in 347 aa overlap:

		10	20	30	40	50	60
10	crf125a.pep	MSGNASSXSSAAIGLWFGAAVSIAEISTGTLLAPLGWQGLAALLLGHAVGGALFFAA					
	crf125-1	MSGNASSPSSSSAIGLWFGAAVSIAEISTGTLLAPLGWQGLAALLLGHAVGGALFFAA					
		10	20	30	40	50	60
15	crf125a.pep	AYIGALTGXKSMESVRLSPGKRGSVLFSVANMLQAGWTAVMYIYAGATVSSALGKVLWDG					
	crf125-1	AYIGALTGRSSMESVRLSPGKRGSVLFSVANMLQAGWTAVMYIYAGATVSSALGKVLWDG					
		70	80	90	100	110	120
20	crf125a.pep	ESFVWVALANGALIVLWLVFGARKTGGLKTVSMLMLLWLVLSAEKFTAGSTAAKXVD					
	crf125-1	ESFVWVALANGALIVLWLVFGARKTGGLKTVSMLMLLWLVLSAEVFFSTAGSTAAQVSD					
		130	140	150	160	170	180
25	crf125a.pep	GMSFGTAVELSAVMPLSWLPLAADYTRHARRPFAATLTATLAYTLTGCCWMYALGLAALF					
	crf125-1	GMSFGTAVELSAVMPLSWLPLAADYTRHARRPFAATLTATLAYTLTGCCWMYALGLAALF					
		190	200	210	220	230	240
30	crf125a.pep	TGETDVAKILLGAGLGAAGILAVVLSTVTTTFLDAYSAGVSANNISAKLS EPIAVAVAV					
	crf125-1	TGETDVAKILLGAGLGAAGILAVVLSTVTTTFLDAYSAGVSANNISAKLS EPIAVAVAV					
		250	260	270	280	290	300
35	crf125a.pep	VGTLLAVLLP VTEYENFLLL IGSVFAPMAA VLIADFFVLK RREEIEG					
	crf125-1	IGTVLAVMLP VTEYENFLLL IGSVFAPMAA VLIADFFVLK RREEIEG FDFAGLVWLAGF					
		310	320	330	340	350	360

Homology with a predicted ORF from *N.gonorrhoeae*

45 ORF125 shows 86.2% identity over a 65aa overlap with a predicted ORF (ORF125ng) from *N.gonorrhoeae*:

	crf125.pep	ACASANNISARFAETPVAVSVTLICTVLAV	30
	crf125ng	KILLGAGLIGITGILAVVLSTVTTTFLDAYSAGASANNISARFAEIPVAVGVTLIRTVLAV	308
50	crf125.pep	MLPVTYENFLLIGSVFAPM-GGFDCLRLKETA	64
	crf125ng	MLPVTYENKFLIRSVFGPMAGGFDCLRLKETA	343

An ORF125ng nucleotide sequence <SEQ ID 805> was predicted to encode a protein having amino acid sequence <SEQ ID 806>:

1 MSGNASSPSS SAAIGLVWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH
 51 AVGGALFFAA AYIGALTGRS SMESVRLSPG KCGSVLFSVA NMLQAGWTA
 101 VMIYAGATVS SALGKVLWDG ESEFVWALAN GALIVLWLVF GARRTGGKLT
 151 VSMILMLLV LMLSSEVFT AGSTAAVVD GMSFGTAVEL SAVMPLSWLP
 201 PLAADYTRHAR RPPAATITAT LAYTLTGCCWM YALGLAALF TGETDVAKIL
 251 LLGAGLGTG ILAVVLSTVTT TFLDAYSAG ASANNISARF AEIPVAVGVTV
 301 LIRTVLAVML PVTYENKFLIR SVFGPMAGGFDCLRLK KTA*

Further work revealed the following gonococcal DNA sequence <SEQ ID 807>:

```

1  ATGTCGGGCA  ATGCTCTCTC  TCTTTCATCT  TCCGCCGCCA  TCGGGCTGGT
51  TTGGTTCGGC  GCGCGCGTAT  CGATTGCCGA  AATCAGACAG  GGATCGCTGC
101  TCGCCOCCCTT  GGGCTGGCAG  CGCGGTCTGG  CGGCCCTGCT  TTTGGGTCACT
151  CGCGTTCGGC  GCGCGCTGTT  TTTTTCGGCG  GCGTATATCG  GCGCACTGAC
201  CGGACGAGC  TCGATCGAAA  GTGTGCGCCT  TCGTTCGGC  AATCGCGGTT
251  CACTGCTGTT  TCCGCGCGCG  APATATCTGC  AACTGGCGCG  CTGACGCGGT
301  GTGATGATT  ACGTCCGCGC  AACGCTCAGC  TCCGCTTTGG  GCAAAGTGTT
351  GTGGGACGGC  GAATCTCTTG  TCTGGTGGCG  ATTGGCAAC  GCGCGACTGA
401  TCGTGCTGTG  GCTGGTTTTC  GCGCGACGCA  GAAOGGCGCG  CTTGAAAACG
451  GTTTCGATGC  TCGTGTATCT  GTTTCGCGTG  TTGTGGTTGA  GCGTCGAAGT
501  GTTCGCTTCG  TCCGGCACAA  ACGCGCGGCC  CGCGCTTTCA  GACGGCATGA
551  CCTTCGGAAC  GGCAGTCGAA  CTGTCCGCGC  TCATGCCGCT  TTCTGGCTGA
601  CGCGTGGCGC  CGCACTACAC  GCGCCACGCA  CGCGCCCGCT  TTGCGGCAAC
651  CCTGACGGCA  ACGCTCGCCT  ATACGCTGAC  GGGCTGCTGG  ATGTATGCTT
701  TGGGTTTGGC  GCGCGCTCTG  TTTACCGGAG  AARCCGACGT  GCGCAAAATC
751  CTGTTGGGCG  GCGGCTTGGG  CATAACGGCG  ATTCTGGCAG  TCGTCTCTCT
801  CACCGTTACC  ACAACGTTTC  TCGATACTTA  TCCGCGCGCG  GCGAGTGGCA
851  ACAACATTTC  CGCGCGTTTT  GCGGAAATAC  CCGTGCCTGT  CGCGGTTACC
901  CTGATCGGCA  CGGTGCTTGC  CGTCATGCTG  CCGCTTACCG  AATATAAAAA
951  CTTCTCGCTG  CTTATCGGCT  CGGTATTTCG  GCGCGATGGC  GCGGTTTTGA
1001  TTGCGGACTT  TTTGCTCTTA  AAGCGCGGTG  AGGAGATTGA  AGGCTTTGAC
1051  TTTGCGGAC  GGTGCTGATG  GCTGCGAGCG  TTTCTCTCT  ACGCTTCTCT
1101  GCTCTCGTCC  GCTTGGGAAA  GCGAGCTCGC  TCTGACGCGC  CCGGTAAATG
1151  CTGCGCTTGC  CATTGCCACC  GTATCGGTAC  GCCTTTCTCT  TAAAAAACCC
1201  CAATCTTAC  AAGGAACCC  GTCATGA

```

This corresponds to the amino acid sequence <SEQ ID 808; ORF125ng-1>:

```

1  MSGNASSPSS  SAAIGLVWFG  AAVSIAEIST  GTLLAPLGWQ  RGLAALLGH
51  AVGGALFFAA  AYIGALTGRS  SMESVRLSFG  KCGSVLFSVA  NMLQLAGWTA
101  VMYIVGATVS  SALGKVLWDG  ESFVWMLAN  GALIVLWLVF  GARRTGGLKT
151  VSMLLMLLAV  LWLSVEVFAS  SGTNAAPAVS  DGMTFGTAVE  LSAMPLSWL
201  PLAADYTRQA  RFFPAATLTA  TLAYTLTGCV  MYALGLAAAL  FTGETDVAKI
251  LLGAGLGITG  ILAVVLSTVT  TTFLDTSYAG  ASANNISARF  AEIPVAVGVT
301  LIGTVLAVML  PVTEYKNFLL  LIGSVFAPMA  AVLIADFFVL  KRREEIEGFD
351  FAGLVWLAG  FILYRFLSS  GWESSIGLTA  FVMSAVALAT  VSVRLFPEKT
401  QSLQRNFS*

```

ORF125ng-1 and ORF125-1 show 95.1% identity in 408 aa overlap:

```

10      20      30      40      50      60
40  orf125-1.pep  MSGNASSPSSSSAIGLVWFGAARVSTAEISTGTLAPLGWQORGLAALLIGHAVGGALFFAA
    orf125ng-1  MSGNASSPSSSSAIGLVWFGAARVSTAEISTGTLAPLGWQORGLAALLIGHAVGGALFFAA
10      20      30      40      50      60
45  orf125-1.pep  AYIGALTGRSSMESVRLSFGKRGSVLFSVANMLQLAGWTAVMIYAGATVVSALGKVLWDG
    orf125ng-1  AYIGALTGRSSMESVRLSFGKCGSVLFSVANMLQLAGWTAVMIYVAGTAVVSALGKVLWDG
70      80      90      100     110     120
50  orf125-1.pep  ESFVWMLANGALIVLWLVFGARRTGGLKTVSMLMLLAVLWLSAEVFSSTAAGTAQ-VS
    orf125ng-1  ESFVWMLANGALIVLWLVFGARRTGGLKTVSMLMLLAVLWLSVEVFSSGTNAAPAVS
130     140     150     160     170     180
55  orf125-1.pep  DGMSFGTAVELSAVMPLSWLPLAADYTRHARRPFAATLTATLAYTLTGCMYALGLAAAL
    orf125ng-1  DGMTFGTAVELSAVMPLSWLPLAADYTRQARRPFAATLTATLAYTLTGCMYALGLAAAL
190     200     210     220     230     240
60  orf125-1.pep  FTGETDVAKILGAGLGAGILAVVLSTVTTTFLDYSAGASANNISARFAETPVAVGVT
    orf125ng-1  FTGETDVAKILGAGLGITGILAVVLSTVTTTFLDYSAGASANNISARFAEIPVAVGVT
250     260     270     280     290     300
65  orf125-1.pep  FTGETDVAKILGAGLGAGILAVVLSTVTTTFLDYSAGASANNISARFAEIPVAVGVT
    orf125ng-1  FTGETDVAKILGAGLGAGILAVVLSTVTTTFLDYSAGASANNISARFAEIPVAVGVT
250     260     270     280     290     300

```

		300	310	320	330	340	350	359
5	orf125-l-pep	LIGTLVLA	VMLPVTEY	ENFLLIG	SVFAPMAA	VLIAEFFV	LKRREEIE	GDFAGLV
	orf125ng-l	LIGTLVLA	VMLPVTEY	KNFLLIG	SVFAPMAA	VLIAEFFV	LKRREEIE	GDFAGLV
		310	320	330	340	350	360	
		360	370	380	390	400		
10	orf125-l-pep	FILYRFL	LSSGWESS	IGLTAPV	MSAVAIAT	VSRLFFK	KTQSLQR	NPSPX
	orf125ng-l	FILYRFL	LSSGWESS	IGLTAPV	MSAVAIAT	VSRLFFK	KTQSLQR	NPSPX
		370	380	390	400			

Based on this analysis, including the presence of putative leader sequence and transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 96

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 809>:

20	1	ATGACCCGTA	TGCGCATCCT	CGCGGGGGCG	CTCTCGGGAA	GGGTGACCGC
	51	GTTCGAGCTT	CGAGAACAG	GTATCAGAT	TGCATTTTC	GATAAAGCT
	101	GCGCCGGGG	CGAACCGCC	CGCCCTATG	TGCGCGCCG	CATGCTCGCG
	151	CCTGCAGCGG	A.ACGGTCGA	AGCCACGCC	GAAGTGCTCA	GGGTGGCGAG
	201	CGAGAGCATC	CGCGTTTGGC	CGCGCATCCG	ATGCCGCTCTG	AACACGCACA
	251	CGATGATGCA	GGAAAACGGC	AGCCTGATTG	TATGGCACGG	CGAGGACAAG
25	301	CCATTATCCA	CGGAGTTCGT	CGCCCATCTC	AAAGCGGGCG	CGGT.ACGGA
	351	TGACGAATC	GTCCGTTGGC	GCGCGACGA	CATCGCGAA	CGCGAACCGC
	401	AACTCGGCGG	ACGTTTTTAA	GACGGCATCT	ACCTGCGCAG	CGAAGC.CAG
	451	CTCGACGGCG	GGCAATATTA	GTCTGCACTT	GCGGACGCTT	TGGACGAAC
	501	GAACGTCGCC	TGCCATTGGG	AACACGAATG	CGTCCCGGAA	GCCTGCAAG..

30 This corresponds to the amino acid sequence <SEQ ID 810; ORF126>:

	1	MTRIAILGGG	LSGRLLALQL	AEQGYQIALF	DKSCRRGEHA	AAYVAAAMLA
	51	PAAXTVFATP	EVVRLGRQSI	PLWRGIRCLR	NHTMMQENG	SLIVWHQDQK
	101	PLSSEFVRHL	KRGXGTDDEI	VWRADDDIAE	REFQLGGREF	DGIYLPTEXQ
	151	LDGRQLXSAL	ADALDELNVF	CHWEHECVPE	ACK...	

35 Further work revealed the complete nucleotide sequence <SEQ ID 811>:

	1	ATGACCCGTA	TGCGCATCCT	CGCGGGGGCG	CTCTCGGGAA	GGGTGACCGC
	51	GTTCGAGCTT	CGAGAACAG	GTATCAGAT	TGCATTTTC	GATAAAGCT
	101	GCGCCGGGG	CGAACCGCC	CGCCCTATG	TGCGCGCCG	CATGCTCGCG
	151	CCTGCAGCGG	AAGCGGTCGA	AGCCACGCC	GAAGTGCTCA	GGGTGGCGAG
40	201	CGAGAGCATC	CGCGTTTGGC	CGCGCATCCG	ATGCCGCTCTG	AACACGCACA
	251	CGATGATGCA	GGAAAACGGC	AGCCTGATTG	TATGGCACGG	CGAGGACAAG
	301	CCATTATCCA	CGGAGTTCGT	CGCCCATCTC	AAAGCGGGCG	CGGTGACGGA
	351	TGACGAATC	GTCCGTTGGC	GCGCGACGA	CATCGCGAA	CGCGAACCGC
	401	AACTCGGCGG	ACGTTTTTCA	GACGGCATCT	ACCTGCGCAG	CGAAGGCCAG
45	451	CTCGACGGCG	GGCAATATTT	GTCTGCACTT	GCGGACGCTT	TGGACGAAC
	501	GAACGTCGCC	TGCCATTGGG	AACACGAATG	CGTCCCGGAA	GGCTGCAAG
	551	CCCAATACGA	CTGGCTGATC	GACTGCGCGG	GCTACGCGCG	AAAAACCGCG
	601	TGGAACCAAT	CCCCCGAGCA	CACCAACACC	CTGCGGGGCA	TACGCGGGCA
	651	AGTGGCGGG	GTTTACACAC	CCGAATACAC	GCTCAACCGC	CCCGTGCGTC
50	701	TGCTCCATCC	GGGTATCCG	CTCTACATCG	CCCGCAAGA	AAACCAACGTC
	751	TTCGTATCG	GCGGACGCCA	AATCGAAAGC	GAAAGCAAG	CCCCCGCAG
	801	CGTGGGTTCA	GGGTGGGAAC	TCTGTTCGCG	ACTCTATGCC	ATCCACCCCG
	851	CTTTCGGGGA	AGCGGACATC	CTCGAAATCG	CCACCGCGCT	GCGGCCACCG
	901	CTCAACCAAC	ACGACCCCGA	AATCGGTTAC	AACGCGCGCC	GACGCTGAT
55	951	TGAATCTAC	GGCTTTTCC	CGCAGAGTTT	CATGATCTCC	CCCGCGTAA
	1001	CGCGCGCGCG	CGCGAGATTG	GCAGTGCGAC	TGTTTGAAGG	AAAGACGCG
	1051	CCCGAACCGG	ATAAAGAAAG	CGGTTTGGCG	TATATCGGAA	GACAAGATTA
	1101	A				

This corresponds to the amino acid sequence <SEQ ID 812; ORF126-1>:

```

1  MTRIAILGGG LSGRLTALQL AEQGYQIALF DKGCRERGEHA AAYVAAAMLA
51  PAEEAVEATP EVVRLGRQSI PLWRGIRCLL NTHTMQENG SLIVWHGQDK
101  PLSSEFVRHL KRGGVADDEI VWRADDIAE REPQLGGRFS DGIYLPTEGQ
151  LDGRQILSAL ADALDELNVP CHWEHECVPE GLQAGYDWLI DCRGYGAKTA
201  WNQSPHTST LRGIRGEVAR VYTPEITLNR FVRLHPFRYP LYIAPKENIV
251  FVIGATQIES ESQAPASVRS GLELLSALYA IHPAFGEADI LEIATGLRST
301  LNHNNFEIRY NRARRLIEIN GLFRHGMIS FAVTAAARL AVALFDGKDA
351  PERDKESGLA YIRRDQ*

```

10 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF126 shows 90.0% identity over a 180aa overlap with an ORF (ORF126a) from strain A of *N. meningitidis*:

```

15  orf126.pep  10      20      30      40      50      60
      MTRIAILGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAAYVAAAMLAPAAXTVEATP
      |||||
orf126a      MTRIAILGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAAYVAAAMLAPAAEAVEATP
      10      20      30      40      50      60

20  orf126.pep  70      80      90      100     110     120
      EVVRLGRQSIPLWRGIRCLNTHTMQENGSLIVWHGQDKPLSSEFVRHLKRGGXTDDEI
      |||||
orf126a      EVVRLGRQSIPLWRGIRCLNTHTMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDDXI
      70      80      90      100     110     120

25  orf126.pep  130     140     150     160     170     180
      VWRADDIAEREPQLGGRFXDGIYLPTEQLDGRQLXSALADALDELNVPCHWEHECVPE
      |||||
orf126a      VWRADDIAEREPQLGGRFXDGIYLPTEQLDGRQLXSALADALDELNVPCHWEHECVPE
      130     140     150     160     170     180

```

The complete length ORF126a nucleotide sequence <SEQ ID 813> is:

```

1  ATGACCGTGA TCGCCATCCT CGGCGCGGGC CTCTCNGGAA GGCTGACCGC
51  ACTGCAGCTT GCAGAAACAAG GTTATCAGAT TGCACTTTTC GATAAAGGCT
101  GCGCGCGGGG CGAACACGCC CGCGCCTATG TTGCGCGCGG CATGCTCGGG
151  CCGTGGCGGG AAGCGGTGCA AGCCACGCGT GAAGTGGTCA GGCTGGGCGA
201  CGAGANCATC CGGCTTTGGC CGGCGATCCG ATGCCATCTG AAAAGCGCTG
251  CCATGATGCA NGAAACGGCG AGCGTGATTG TGTGGCGCGG CGAGGACAAA
301  CCTTTATCCA ACGACGTCCT CGCGCATCTC AAGACGGCGG GCGTAGCGGA
351  TGCANAAATC GTCCGTGTGGC CGCGCAGAGA CATCGCGGAA CGGCAACCGC
401  AACTCGGCGG ACCTTTTTC AAGCGCATCT ACCTGCGGAG CAGAGCCGAG
451  CTCGACGCGG GGCAAATATT GTCTGCACCT GCGACGCGCT TGGAGCAACT
501  GAACGTCCCC TGCCATTGGG AACACGAATG TGCCCGCGAA GACTTGCAAG
551  CCCAATAAGA CTGGCTGATC GACTGCGCGG GCTACGCGCG AAAAACCGGG
601  TGGAAACCAAT CCCCCGANN AACCAGCAAC CTCGCGGGCA TACGGGCGGA
651  AGTGGCGCGG GTTACACAC CGGAATCAC GCTCAACCGG CCGCTGCGCC
701  TGCTACACCC CGGCTATCCG CTNTACATCG CCCCAGAAAG AAACNCGCT
751  TTCTGTCATCG GCGCGACCCA AATCGAAAGC GARAGCCAGC CACCTGCCAG
801  CGTGGCTTCC GGGCTGGAAC TCTTATCCCG ACTCTATGCC GTCCACCCCG
851  CCTTGGCGGA AGCGGACATC CTCGAAATCG CCACGCGGCT GCGGCCACAG
901  CTCAATCACC ACAACCCCGA AATCGTGTAC AACCGGCGCT GACGCTGAT
951  TGAATCAACG GGCCTTTTTC GGCACGCTTT CATGATCTCC CCGCGCGTAA
1001  CCGCGCGCGC GCTCAGATTG CGATGCGCAC TGTGTTGACG AAAAGANGCG
1051  CCGGACGCGG ATGAAGAAAG CGGTTTGGCG TATATCGGAA GACAGAGATTA
1101  A

```

55 This encodes a protein having amino acid sequence <SEQ ID 814>:

```

1  MTRIAILGGG LSGRLTALQL AEQGYQIALF DKGCRERGEHA AAYVAAAMLA
51  PAEEAVEATP EVVRLGRQSI PLWRGIRCLL NTHTMQENG SLIVWHGQDK
101  PLSSEFVRHL KRGGVADDDI VWRADDIAE REPQLGGRFS DGIYLPTEGQ
151  LDGRQILSAL ADALDELNVP CHWEHECAPE DLQAGYDWLI DCRGYGAKTA
201  WNQSPXHTST LRGIRGEVAR VYTPEITLNR FVRLHPFRYP LYIAPKENIV

```

```

251 FVIGATQIES ESQAPASVRS GLELLSALYA VHPAFGEADI LEIATGLRPT
301 LNHHNPEIYR NRARRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKXA
351 PERDEESGLA YIRRQD*

```

ORF126a and ORF126-1 show 95.4% identity in 366 aa overlap:

```

5      orf126a.pep      MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAAAEVATP      60
      orf126-1          MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAAAEVATP      60
10      orf126a.pep      EVVRLGRQXIPLWRGIRCLNTHMMQENGSLIVWHGQDKPLSSEFVRLHKGKGVADDEI      120
      orf126-1          EVVRLGRQXIPLWRGIRCLNTHMMQENGSLIVWHGQDKPLSSEFVRLHKGKGVADDEI      120
15      orf126a.pep      VWRADDIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHEWECAVE      180
      orf126-1          VWRADDIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHEWECAVE      180
20      orf126a.pep      DLQAQYDWLIDCRGYGAKTAWNQSPXXTSLRGIRGEVARVYTPETITLNRPRVLLHPRYP      240
      orf126-1          GLQAQYDWLIDCRGYGAKTAWNQSPXXTSLRGIRGEVARVYTPETITLNRPRVLLHPRYP      240
25      orf126a.pep      LYIAPKENHVFVIGATQIES ESQAPASVRS GLELLSALYA VHPAFGEADI LEIATGLRPT      300
      orf126-1          LYIAPKENHVFVIGATQIES ESQAPASVRS GLELLSALYA VHPAFGEADI LEIATGLRPT      300
30      orf126a.pep      LNHHNPEIYR NRARRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKXA PERDEESGLA      360
      orf126-1          LNHHNPEIYR NRARRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKXA PERDEESGLA      360
35      orf126a.pep      YIRRQDX
      orf126-1          YIRRQDX
40      orf126a.pep      YIRRQDX
      orf126-1          YIRRQDX
45

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF126 shows 90% identity over a 180 aa overlap with a predicted ORF (ORF126ng) from *N.gonorrhoeae*:

```

50      orf126.pep      MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAAAEVATP      60
      orf126ng          MTRIAVLGGGLSGRLTALQLAEQGYQIELFDKGTQGEHAAAYVAAAMLAPAAAEVATP      60
      orf126.pep      EVVRLGRQXIPLWRGIRCLNTHMMQENGSLIVWHGQDKPLSSEFVRLHKGKGVADDEI      120
      orf126ng          EVVRLGRQXIPLWRGIRCLNTHMMQENGSLIVWHGQDKPLSSEFVRLHKGKGVADDEI      120
      orf126.pep      VWRADDIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHEWECAVE      180
      orf126ng          VWRADDIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHEWECAVE      180

```

60 An ORF126ng nucleotide sequence <SEQ ID 815> was predicted to encode a protein having amino acid sequence <SEQ ID 816>:

1 MTRIAVLGGGLSGRLTALQLAEQGYQIELFDKGTQGEHAAAYVAAAMLAPAAAEVATP

-449-

51 PRAEAVEATP EVIRLGROSI PLWRGIRCLR NTLTMMQENG SLIVWHGQDK
 101 PLSEFVRHL KRGGVADLEI VWRADIAE REPQLGGRFS DGIVLPTGQ
 151 LDGRQLSAL ADALDELNVF CWHHECAPO DLQAOYDWVI DCRGYGAKTA
 201 WNQSPHST LRGIRGEVRG FTRPKSRSTA PCACCTRAIR STSPRKTKTS
 251 SSSARPKSKA KAKPPPAYVP GWNSYPRMSP STPPSAKPTS SKWRPLGRPT
 301 LNHNPEIRY SRERRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKDA
 351 PERDEESGLA YIGRQD*

Further work revealed the following gonococcal DNA sequence <SEQ ID 817>:

1 ATGACCCGTA TCGCGTCTCT CGGAGGCGGC CTTTCGGAA GGCTGACCGC
 10 51 ATTGCAGCTT GCAGAACAA GTTATCAGAT TGAACTTTTC GACAAGGGCA
 101 CCGCCCAAGG CGAACACGCC GCGCGCTATG TTGCGCGCGC GATGCTGCGC
 151 OCTGCGGGCG AAGCGGTGCA GGCAACGCC GAAGTCATCA GGCTGGGCGC
 201 CGAGAGCATT CGCTTTTGGC GCGGCATCCG ATGCGGCTGT AACACGCTCA
 251 CGATGATGCA GGAACACGCC AGCCTGATTG TGTGGCAOGG CGAGGACAA
 15 301 CCATTATCCA GCGAGTTGCT CGCCCATCTC AAGCGCGGCG GCGTGAOGGA
 351 TGACGAAATC GTCGGTTGGC GCGCGATGA AATCGCGGAA CGCGAACCGC
 401 AACTCGGCGG ACGTTTTTCA GACGCGATCT ACCTGCGCAG CGAGAGCCAG
 451 CTGACGCGGC GCGCAATATT GTCTCGACTT GCGACGCTT TGGACGAAT
 501 GACGCTCCCT TCGCTATGCG AACACGAAAG CGCCCCCAAA GACCTGCAR
 20 551 CCGAATACGA CTGGGTAACT GACTGCGGCG GTACGCGGCG GAAACGCGC
 601 TGGAACCAAT CCGCCGAGCA CACGAGCACC TTGCGCGGCA TACGCGGCGA
 651 AGTGCGCGCG GTTTACACGC CGGAATCAC GCTCAACGCG CCGTGCGGCC
 701 TGTGCGACCC GCGCTATCCG CTCTACATCG CCGCGAAGA ARACACGTC
 751 TTGCTCATCG GCGCGACCCA AATCGAAGCG GAAAGCCAAG CCGCCCGCAG
 25 801 CGTACGTTCC GGGGTGGAAC TCTTATCGCG GCTCTATGCC GTCCACCCCG
 851 CTTTGGGCGA AGCGGACATC CTCGAAATCG CCGCGCGGCT GCGCCCGCAG
 901 CTCAACCAAC ACAACCCCGA AATCGGCTAC AGCGCGAAG CGCGGCTCAT
 951 CGAAATCAAC GGCCTTTTCC GGCACGGCTT TATGATTTCC CCGCGCGTAA
 1001 CGCGCGCGCG CGTCAGATTG GCAGTGGCAC TGTGTTGACG AAAAGACGCG
 30 1051 CCGCAACGCTG ATGAGAAAG CGGTTTGGCG TATATCGGAA GACAAGATTA
 1101 A

This corresponds to the amino acid sequence <SEQ ID 818; ORF126ng-1>:

1 MTRIAVLGGG LSGRLTALQ AEOGYQIELF DKGTRQGEHA AAYVAAMLA
 51 PRAEAVEATP EVIRLGROSI PLWRGIRCLR NTLTMMQENG SLIVWHGQDK
 35 101 PLSEFVRHL KRGGVADLEI VWRADIAE REPQLGGRFS DGIVLPTGQ
 151 LDGRQLSAL ADALDELNVF CWHHECAPO DLQAOYDWVI DCRGYGAKTA
 201 WNQSPHST LRGIRGEVR VYTPETLNAR PVRLLRHRYV LYIAPKENRV
 251 FVIGATQIES ESQAPASVRS GLELLSALYA VHPAFGEADI LEIAAGLRPT
 40 301 LNHNPEIRY SRERRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKDA
 351 PERDEESGLA YIGRQD*

ORF126ng-1 and ORF126-1 show 95.1% identity in 366 aa overlap:

		10	20	30	40	50	60
crf126-1.pep	MTRIAVLGGGLSGRLTALQ	LAEOGYQIALFDK	GCGRRGEHAAAYVAAMLA	PRAEAVEATP			
45 crf126ng-1	MTRIAVLGGGLSGRLTALQ	LAEOGYQIELFDK	GTGTRQGEHAAAYVAAMLA	PRAEAVEATP			
	10	20	30	40	50	60	
	70	80	90	100	110	120	
50 crf126-1.pep	EVIRLGROSIPLWRGIRCLR	NTLTMMQENGSLIVWHGQDK	PLSEFVRHLKRGGVADLEI				
crf126ng-1	EVIRLGROSIPLWRGIRCLR	NTLTMMQENGSLIVWHGQDK	PLSEFVRHLKRGGVADLEI				
	70	80	90	100	110	120	
	130	140	150	160	170	180	
55 crf126-1.pep	VWRADIAEREPQLGGRFS	DGIVLPTGQDLGRQILSALADAL	DELNVFCWHHECAPO				
crf126ng-1	VWRADIAEREPQLGGRFS	DGIVLPTGQDLGRQILSALADAL	DELNVFCWHHECAPO				
	130	140	150	160	170	180	
	190	200	210	220	230	240	
60 crf126-1.pep	GLQAOYDWLIDCRGYGAKT	AWNQSPHSTLRGIRGEVAR	VYTPETLNARPVRLLRHRYP				
crf126ng-1	DLQAOYDWVIDCRGYGAKT	AWNQSPHSTLRGIRGEVAR	VYTPETLNARPVRLLRHRYP				
	190	200	210	220	230	240	

65

-450-

		250	260	270	280	290	300
	orf126-1.pep	LYIAPKENHVFFVIGATQIESESQAPASVRSGLLELLSALYAHPAFGEADILEIATGRLPT					
5	orf126ng-1	LYIAPKENHVFFVIGATQIESESQAPASVRSGLLELLSALYAHPAFGEADILEIATGRLPT					
		250	260	270	280	290	300
	orf126-1.pep	LNHNHPEIYRNRRRLIEINGLFRHGFMI SPAVTAAARLAVAFDGDADPERDKESGLA					
10	orf126ng-1	LNHNHPEIYRNRRRLIEINGLFRHGFMI SPAVTAAARLAVAFDGDADPERDKESGLA					
		310	320	330	340	350	360
15	orf126-1.pep	YIRRDQX					
	orf126ng-1	YIGRDQX					

Furthermore, ORF126ng-1 shows homology to a putative *Rhizobium* oxidase flavoprotein:

	gi 2627327 (AF004408) putative amino acid oxidase flavoprotein [Rhizobium etli]
	Length = 327
20	Score = 169 bits (423), Expect = 3e-41
	Identities = 112/329 (34%), Positives = 163/329 (49%), Gaps = 25/329 (7%)
	Query: 3 RI AVLGGGLSGRLTALQLAEQGYIELFDKGTROGEHXXXXXXXXXXXXXXXXXXXXX 62
25	Sbjct: 2 RILVNGAGVAGLTVAWQLYRHGFRVTLAERAGTVGA-GASGFAGGMLPAWPCERESAEFP 60
	Query: 63 IRLGRQSIPLWRGIRCLNLTMMQENGSLIVWHGQDKLSSEFVRHLKRGVADDEIVR 122
30	Sbjct: 61 LTLGRLAADWEEA-----LPGHVHRRGTLVAVAGGRDTGELDRFSRRTS-GWEVLDEVA- 113
	Query: 123 WRADIEAEREPLQGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHEHCAQPD 182
35	Sbjct: 114 -----IAALEPDLAGRFRRALFFRQEAHLDFRQALAAAGLEDARMLATLG---VVGES 165
	Query: 183 QAQYDWIDRCRGYAKTAWNQSPHSTLRGIRGEVAVRYTPEITLNRPVRLHPRYPLY 242
40	Sbjct: 166 DVDHDRVVDCTGAA-----QIGRLPGLRGVRGEMLCVETTEVSLSPVRLHPRHIY 218
	Query: 243 IAPKENHVFFVIGATQIESESQAPASVRSGLLELLSALYAHPAFGEADILEIATGRLPTLN 302
45	Sbjct: 219 IVPDRKNRFMVGATMIESDDGGPITARSIMELLNAAAHMHPAFGEARVTETGAGVRFPAYP 278
	Query: 303 HNHPEIYRNRRRLIEINGLFRHGFMI SP 331
	Sbjct: 279 DNLP--RVTOEGRTLHVNGLYRHGFLAP 305

This analysis suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 97

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID

50	819>:	
		1 ATGACTGATA ATCGGGGCTT TAGCTGGT GAATTAATAT CAGTGGCTTT
		51 GATATPCTCT GATCTGCTT TAATGTGTTA TCCGAGCTAT CCGATATATG
		101 TTGAGAAAGC AAGATAAAT CCAATGCGCG CAGCTTGTGT AGAAATGCA
55		151 CATTTTATGG AAAAGTTTAT TCTGCAGAT GGGAGGTTTA AACAAATCT
		201 TACCAAGTGG CCAAGTTTTC CGATTAAAGA GCGCAAGGC TTTGTATCC
		251 GTTTGAATGG AATCGTCGCG CGG. GCTT TAGACAGTAA ATTCATGTTG
		301 AAGGCGGTAG CCATAGATAA AGATAAAAT CCTTTTATTA TTAAGATGAA
		351 TGAAATCTA GTAACCTTTA gTTTGCAAGA AGTCCGCGAG TTCGTGTAGT
		401 GACGCGCTGG ATTATTTTAA AGGAATAGT AAGGACGCA AGTTACTTTA
60		451 GTAG

This corresponds to the amino acid sequence <SEQ ID 820; ORF127>:

```

1 MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLENA
51 HMEKEFYLN GRFKQTSTKW PSLPIKEAEG FCIRLINGIVA RXALDSKFML
101 KAVAIKDKNP FFIKMNENLV TFICKKSASS CSDGLDYFK GNDKCKLLK
151 *

```

Further work revealed the following DNA sequence <SEQ ID 821>:

```

1 ATGACTGATA ATCGGGGGTT TAGCTGGTT GAATTAATAT CAGTGGTCTT
51 GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
101 TTGAGAAAGC AAGATAAAT CAGTCGGGG CAGCCTGTGT AGAAATGCA
151 CATTTTATGG AAAAGTTTTC TCTGCAGAAAT GGGAGGTTTA AACAAACATC
201 TACCAAGTGG CCAAGTTTGC CGATTAAAGA GGCAGAAAGC TTTGTATCC
251 GTTTGAATGG AATCGCGGCG GGGGCTTTAG ACAGTAAATT CATGTTGAAG
301 GCGGTAGCCA TAGATAAAGA TAAAAATCCT TTTATTATTA AGATGAATGA
351 AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTCG TGTAGTGACG
15 GCGTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG

```

This corresponds to the amino acid sequence <SEQ ID 822; ORF127-1>:

```

1 MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLENA
51 HMEKEFYLN GRFKQTSTKW PSLPIKEAEG FCIRLINGIAR GALDSKFMLK
101 AVAIKDKNP FFIKMNENLV TFICKKSASS CSDGLDYFK GNDKCKLLK*

```

20 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF127 shows 98.0% identity over a 150aa overlap with an ORF (ORF127a) from strain A of *N.*

meningitidis:

```

25 orf127.pep 10 20 30 40 50 60
MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLENA HMEKEFYLN
orf127a 10 20 30 40 50 60
MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLENA HMEKEFYLN

30 orf127.pep 70 80 90 100 110 120
GRFKQTSTKW PSLPIKEAEG FCIRLINGIAR XALDSKFMLKAVAIKDKNP FFIKMNENLV
orf127a 70 80 90 100 110 120
GRFKQTSTKW PSLPIKEAEG FCIRLINGIAR XALDSKFMLKAVAIKDKNP FFIKMNENLV

35 orf127.pep 130 140 150
VTFICKKSASS CSDGLDYFK GNDKCKLLK
orf127a 120 130 140 150
VTFICKKSASS CSDGLDYFK GNDKCKLLK

```

The complete length ORF127a nucleotide sequence <SEQ ID 823> is:

```

1 ATGACTGATA ATCGGGGGTT TAGCTGGTT GAATTAATAT CAGTGGTCTT
45 51 GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
101 TTGAGAAAGC AAGATAAAT CAGTCGGGG CAGCCTGTGT AGAAATGCA
151 CATTTTATGG AAAAGTTTTC TCTGCAGAAAT GGGAGGTTTA AACAAACATC
201 TACCAAAATGG CCAAGTTTGC CGATTAAAGA GGCAGAAAGC TTTGTATCC
251 GTTTGAATGG AATCGCGGCG GGGGCTTTAG ACAGTAAATT CATGTTGAAG
301 GCGGTAGCCA TAGATAAAGA TAAAAATCCT TTTATTATTA AGATGAATGA
351 AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTCG TGTAGTGACG
50 GCGTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG

```

This encodes a protein having amino acid sequence <SEQ ID 824>:

```

1 MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN TVRAALLENA
51 HMEKEFYLN GRFKQTSTKW PSLPIKEAEG FCIRLINGIAR GALDSKFMLK
101 AVAIKDKNP FFIKMNENLV TFICKKSASS CSDGLDYFK GNDKCKLLK*

```

ORF127a and ORF127-1 show 99.3% identity in 149 aa overlap:

		10	20	30	40	50	60
5	orf127a.pep	MTDNRGFTLV	ELISVVLILSVLALIV	PSYRNYVEKAKINAVRAALLEN	NAHFMEK	FYLN	
	orf127-1	MTDNRGFTLV	ELISVVLILSVLALIV	PSYRNYVEKAKINAVRAALLEN	NAHFMEK	FYLN	
		10	20	30	40	50	60
		70	80	90	100	110	120
10	orf127a.pep	GRFKQTSTKW	PSLPIKEAEGFCIRL	NGIARGALDSKF	MLKAVAIDKDKN	PFI	IKMNNENLV
	orf127-1	GRFKQTSTKW	PSLPIKEAEGFCIRL	NGIARGALDSKF	MLKAVAIDKDKN	PFI	IKMNNENLV
		70	80	90	100	110	120
		130	140	150			
15	orf127a.pep	TFICKKSASS	CSDDLDFYFGNDK	CKLLK			
	orf127-1	TFICKKSASS	CSDDLDFYFGNDK	CKLLK			
		130	140	150			

20 Homology with a predicted ORF from *N.gonorrhoeae*

ORF127 shows 97.3% identity over a 150 aa overlap with a predicted ORF (ORF127ng) from

N.gonorrhoeae:

	orf127.pep	MTDNRGFTLV	ELISVVLILSVLALIV	PSYRNYVEKAKINAVRAALLEN	NAHFMEK	FYLN	60
25	orf127ng	MTDNRGFTLV	ELISVVLILSVLALIV	PSYRNYVEKAKINAVRAAFL	ENAHFMEK	FYLN	60
	orf127.pep	GRFKQTSTKW	PSLPIKEAEGFCIRL	NGIARGALDSKF	MLKAVAIDKDKN	PFI	IKMNNENLV
	orf127ng	GRFKQTSTKW	PSLPIKEAEGFCIRL	NGIARGALDSKF	MLKAVAIDKDKN	PFI	IKMNNENLV
30	orf127.pep	VT	FICKKSASSCSDDLDFYFGNDK	CKLLK	150		
	orf127ng	VT	FICKKSASSCSDDLDFYFGNDK	CKLLK	149		

The complete length ORF127ng nucleotide sequence <SEQ ID 825> is:

35	1	ATGACTGATA	ATCGGGGTT	TACACTGGTT	GAATTAATAT	CAGTGGCTTT	
	51	GATATTGCTC	TGACTTGCCT	TAATTGTTA	TCCGAGCTAT	CCGATATG	
	101	TTGAGAAAGC	AAGATAAAT	GCATCGGG	CAGCCTTGT	AGAAATGCA	
	151	CATTTTATGG	AAAAGTTT	TCTGCGAAT	GGGAGATT	AACAACATC	
	201	TACCAATAGG	CCAAGTTTGC	CGATTAAAGA	GCAGAAAGC	TTTGTATCC	
40	251	GTTTGAATGG	AATCGCGCGC	GGGGCTTTAG	ACAGATAAAT	CATGTTGAAG	
	301	CGCGTAGCCA	TAGATAAAGA	TAAAAATCCT	TTTATTATTA	AGATGAATGA	
	351	AAATCTAGTA	ACCTTTATTT	GCAAGAAGTC	CGCCAGTTCG	TGTATGACG	
	401	GGCTGGATTA	TTTAAAGGA	AATGATAAGG	ACTGCAAGT	ACTTAAGTAG	

This encodes a protein having amino acid sequence <SEQ ID 826>:

45	1	MTDNRGFTLV	ELISVVLILSVLALIV	PSYRNYVEKAKINAVRAAFL	ENAHFMEK	FYLN	
	51	HFMEK	FYLN	GRFKQTSTKW	PSLPIKEAEG	FCIRLNGIAR	GALDSKFMLK
	101	AVAIDKDKNP	FIIKNNENLV	TFICKKSASS	CSDDLDFYFG	NDKDKLLK*	

ORF127ng and ORF127-1 show 100.0% identity in 149 aa overlap:

		10	20	30	40	50	60
50	orf127-1.pep	MTDNRGFTLV	ELISVVLILSVLALIV	PSYRNYVEKAKINAVRAALLEN	NAHFMEK	FYLN	
	orf127ng-1	MTDNRGFTLV	ELISVVLILSVLALIV	PSYRNYVEKAKINAVRAALLEN	NAHFMEK	FYLN	
		10	20	30	40	50	60
		70	80	90	100	110	120
55	orf127-1.pep	GRFKQTSTKW	PSLPIKEAEGFCIRL	NGIARGALDSKF	MLKAVAIDKDKN	PFI	IKMNNENLV
	orf127ng-1	GRFKQTSTKW	PSLPIKEAEGFCIRL	NGIARGALDSKF	MLKAVAIDKDKN	PFI	IKMNNENLV
		70	80	90	100	110	120

-453-

```

              130      140      150
orfl27-1.pep  TFICKKSASSCSGDLDFYFGNDKDKCKLLKX
              |||
5  orfl27hg-1  TFICKKSASSCSGDLDFYFGNDKDKCKLLKX
              130      140      150

```

This analysis, including the fact that the predicted transmembrane domain is shared by the meningococcal and gonococcal proteins, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

10 Example 98

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 827>

```

1  ..GTGCTGCTGG CTTCGGTGAT TGCTCTCAA ATCTTCCTTT ACGAAGATT
51 CAACCAAATG CGGAAACCG GTGGAGCAT CTGCGGTTTT CTGTCCAAAT
101 ATTTATCTGG GTTTTCAGCA GGGGTATTTC GATTTGAGTG CCGACGAGAA
151 CCCGCTACTG CATATCTGGT CTTTGGCAGT AGAGGAACAG TATTACCTCC
201 TGATATCCCT TTTGCTGATA TTTTGTGCA AAAAAACCAA ATCGCTACGG
251 GTGCTGGGTA ACATCAGCAT CATCCTGTT TTGATTTTGA CTGCCTCATC
301 GTTTTTGCCA AGCGGGTTTT ATACCGACAT CCTCAACCAA CCAATATCT
351 ATTACTCTTC GACACTGAGG TTTCCGAGC TGTGGCAGG TTGCTGTGCT
20 401 GCGGTTTACG GGCACACGCA AAACGGCAGA CGGCAACAG CAAATGGAAA
451 ACGCGAGTTG CTTTCAICAC CTGTCTCGG CGCATGCTTT GCCTGCCTGT
501 TCGTATGCA CAAACACAT CCGTATATCC CGGAATGAC CTTGCTCCTT
551 CCCTGCTGCG TGACGGCACT GCTTATCCGG AGTATGCAAT ACGGGACACT
601 TCCGACCCGC ATCCTGTGCG CAAGCCCAT CGTATTTGTC GGCAAAATCT
25 651 CTTATTCCCT ATACCTGTAC CATTGGATT TTATTGCTTT GCCTCCGCTC
701 ATTAGAGGCG GGAACACGCT CGACTGCCT CGCG.

```

This corresponds to the amino acid sequence <SEQ ID 828; ORF128>:

```

1  ..VSLASVIAQ IFLYEDFNQM RKTVELSAVF LSNLYLFQO GYFDLSADEN
51 PVIRHISLAV EEQYLLYPL LLIPOCKKTK SIRVLRNISI ILFLILTASS
30 101 FLPSGFYTDI LNPNTYLYL LTRFPELLAG SLLAVYQQT NGRRTTANGK
151 RQLSSSLFCG ALLACLFLVID KHNPFIPGMT LLLPCLLTAL LIRSMQYGTL
201 PTRILSASPI VFVGKISYSL YLHNWIFIAF APLIRGGKQL GLPA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 829>:

```

1  ATGCAAGCTG TCCGATACAG ACCGGAATTT GACGGAATTG GGGCCGTCGC
35 51 CGTGCATATCC GTCATGATTT TCCACTGAA TAACCGCTGG CTGCCCGGAG
101 GATTCTCTGG GGTGGACATT TTCTTTGTCA TCTCAGGATT CCTCATTAAC
151 GGCACTCATTC TTTCTGAATT ACAGAAOOGT TCTTTTCTTT TCGGGGATTT
201 TTATACCCCG AGGATTAAAG GGAATTAATC TGCTTTTATT GCGGCGSTGT
40 251 CGCTGGCTTC GGTGATGTCC TCTCAATATC TCTTTTACGA AGATTTCAAC
301 CAAATGCGGA AAACCGTGGG GCTTTCTGCG GTTTTCTGTT CCAATATTTA
351 TCTGGGTTTT CAGCAGCGCT AATTCCGATT GAGTCGCGAG CGAACCCCGG
401 TACTGCAATT CTGCTCTTTC CAGACTATTA AGACTATTA CTTCTCTTAT
451 CCCCTTTTGC TGATATTTTC TCGCAAAAA ACCAATATGAC TACGGGTGCT
501 GCGTAAATAT AGCATCATCC TGTTTTGTAT TTTGACTGCC TCATCTTTTT
45 551 TGCCGAAGCG GTTTTATACC GACATCCTCA ACCAACCCAA TACTTATATC
601 CTTTCGACAC TGAGGTTTCC CGAGCTGTGG CAGAGTTCGG TGCTGGCGGT
651 TTACGGGCAA ACCGAAACAG CGACAGCGCA AACAGCAAA GTGAAACCGCG
701 AGTTGCTTTC ATCACTCTGC TCTGGGCAAT TGCTTGCTCG CTTGTTCTGT
751 ATTGACAAAC ACAATCCGTT TATCCGCGGA ATGACCCCTG TCTTTCCCTG
50 801 CTTGCTGACG GCACTGCTTA TCCGGAGTAT GCAATACGGG ACACCTCCGA
851 CCGCATCTCT GTGCGCAAGC CCGATCGTAT TTGTCGGCAA AATCTCTTAT
901 TCCCTATACC TGTACCATTT GATTTTATTT GCTTTGCCCC ATTACATTAC
951 AGGCGACAAA CAGCTCGGAC TGCTGCGGT ATCGCGGTTT GCGCGTTGA
100 1001 GCGCGGATTT TTCCCTGTTG AGTTTATATT TGATTTGAAC CGCGCTTAGA
1051 AAACGGGAAGA TGACCTTCAA AAAGGCATTT TTCTGCTCTT ATCTCGCCCC
1101 GTCCCTGATA CTTTGTGCTT ACAACCTGTA CGCAAGGGGG ATATTGAAC
1151 AGGAACACAT CCGCCCGTTG CCGCGCGGCG CCGTGTCTGC GGAAATCAT

```


-455-

5	orf128a	ILFLILITATSF	LPSPGFYTD	ILNQNTYYL	SLTRFP	ELLAGS	LLAVYGTQ	TNGRRQT	ANGK
		180	190	200	210	220	230		
10	orf128.pep	RQLSSSL	CGFALL	ACLV	FDKHN	PFPGMT	LLPCLL	TALLIR	SQMYGTL
		160	170	180	190	200	210		
15	orf128a	RQLSSSL	CGFALL	ACLV	FDKHN	PFPGMT	LLPCLL	TALLIR	SQMYGTL
		240	250	260	270	280	290		
20	orf128.pep	VFVGKIS	YSLY	YHWF	FIAFA	PLIRG	GKQL	GLPLA	
		220	230	240					
25	orf128a	VFVGKIS	YSLY	YHWF	FIAFA	HYITG	DKQL	GLPAV	SAVAAL
		300	310	320	330	340	350		
30	orf128a	KMTFKAF	FCILY	LAPSL	ILVGN	LYARG	ILKQ	EHLR	PLPGAP
		360	370	380	390	400	410		

The complete length ORF128a nucleotide sequence <SEQ ID 831> is:

20	1	ATGCAAGCTG	TCCGATACAG	ACCGGAAAT	GACGGATTGC	GGGCGTGC	CGC
	51	CGTGCATACC	GTCATGATT	TCCACCTGAA	TAAACCGCTG	CTGCCGGAG	CGC
25	101	GATTCTGGG	GGTGGACATT	TTCTTTGTCA	TCTCAGGATT	CCTCATTACC	
	151	GGCATCATT	TTCTGAAAT	ACAGAACGGT	TCTTTTCTT	TCCGGGATT	
30	201	TTATACCCGC	AGGATTAAAG	GGATTATCC	TGCTTTTATT	CGCGCGTGT	
	251	CGCTGGCTTC	GGTATTGCC	TCTCAATCT	TCTTTTAGA	AGATTTCAC	
35	301	CAATGCGGA	AAACCGTGA	GCTTCTGCG	GTTTCTGT	CCAATTTTA	
	351	TCTGGGGTT	CAGCAGGGT	ATTTCGATT	GAGTCCGAC	GAGAACCCG	
40	401	TACTGCATAT	CTGGTCTTG	GCAGTAGAG	AACAGTATTA	CCTCTGTAT	
	451	CCTCTTTTG	TGATATTTG	CTGCAAAAA	ACAAAATGC	TACGGGTGT	
45	501	GCGTAACAT	AGCATCATC	TATTTCTGAT	TTTGACTGCC	ACATCGTTT	
	551	TGCCAAGCG	GTTTATACC	GATATCTCA	ACCAACCCA	TACTATTAC	
50	601	CTTTGACAC	TGAGGTTTCC	CGAGCTGTG	CGAGTGTGC	TGCTGGCGT	
	651	TTACGGGCA	ACAGCAAGC	GACGAGGCA	ACAGCAAAAT	GGAAACGCG	
55	701	AGTTCGTTT	ATCAGCTGC	TTCCGGCAT	TCTTTCCTG	CCCTTCCTG	
	751	ATTGACAAAC	ACAAATCCG	TATCCCGGA	ATGACCTGC	TCTTCCCTG	
60	801	CCCTGTCAG	CACTGCTTA	TCCCGGATAT	GCAATACGG	ACACTCCGA	
	851	CCCGCATCT	GTCGGCAAG	CCCATCGTAT	TTGTCGGCA	AATCTCTAT	
65	901	TCCCTATACC	TGACCATATG	GATTTTATT	GCTTTCGCC	ATTACATTAC	
	951	AGGCGCAAA	CAGCTGGAC	TGCTCGCGT	ATCGGCGGT	GCCGCGTTA	
70	1001	CGGCGGATT	TTCCCTGTT	AGTTATTAT	TGATTGAACA	GCCGCTTGA	
	1051	AAACGGAGA	TGACCTTCA	AAAGGCATT	TTCTCGCTCT	ATCTCGGCC	
75	1101	GTCCTGATA	CTTGTGGTT	ACAACCTGA	CGCAAGGGG	ATATTGAAC	
	1151	AGGAACACT	CCGCCCGTT	CCCGCGCGC	CCCTTGTGC	GGAAAAATC	
80	1201	TTTCCGAAA	CCGTCCTGAC	CTCGGCGAC	TGCGCAGCG	GACACCTCG	
	1251	GGGGTTCTG	GATTATTCG	GCAAGCGGA	AGGGTGGAA	GCCAAAAATC	
85	1301	TGTCCTCGA	TTCCGAGTGT	TGGTTTGGG	TAGATAGAA	GCTGCGCAG	
	1351	AACCCGTTAT	GTGCAAAAA	CCGGGATGAA	GTGGAAGA	CCGAACCGT	
90	1401	TTTCATTGCC	CAATCTTATG	ATTGAGGAT	GGGCGGCCG	CCGCTGCGA	
	1451	GATTTGAGG	GCAATCTCT	CTAATACCG	GCTTCCGAC	CGATTCAGG	
95	1501	GAAACCGCA	AAAGGATAG	CGCCCTCAA	CCCTCTAG	TTTTCGAAA	
	1551	CACACATCA	ATCAGCGTT	CCCGCTGAG	GGAGAAAAA	TGAAAGAGT	
100	1601	TTCCGCAAA	CCAAATCTC	CCCCCATTC	AGGCTGGG	CGACATCGC	
	1651	AAGAGCAAT	AGGGGCTCT	TGATTGTAT	AAAGATATTC	CCAATGTCA	
105	1701	TTGGGTGGAC	GCACAAAAAT	ACCTGCCAA	AAACACGTC	GAAATATAC	
	1751	GCGCTATCT	TTACGGCGAC	CAAGACACC	TGACCTATT	CGGTTCTTAT	
110	1801	TATATGGGG	GGGAATTCA	CAAAACGGA	CGCCTGTTA	AATCTTCTG	
	1851	CGACGGGCA	TTGCAGTAG				

This encodes a protein having amino acid sequence <SEQ ID 832>:

60	1	MQAVRYRPEI	DGLRAVAVL	VMIFLNNRW	LPGGFLVDI	FFVISGFLIT	
	51	GILLSEIQNG	SFSFRDFYTR	RIKRIYPAFI	AAVSLASVIA	SOIFLYEDFN	
65	101	QMRKTVELSA	VFLSNIYLF	QGGYFDLSAD	ENPVLHWSL	AVEEQYLLY	
	151	PLLLIFCCCK	TKSLRLRNI	SIILFLILTA	TSFLESPGYT	DILNQNTYY	
70	201	SLTRFPPELL	AGSLAVYGT	TNGRRQTAN	GKROLLSSIC	FGALLACL	
	251	IDKHNFFIPG	MTLLLPCLLT	ALLIRMSQY	TLPRILSAS	PIVFGKISY	
75	301	SLYLYHWIFI	AFAHYITGDK	QLGLPAVSA	AALTAGFSL	SYYLIEQPLR	
	351	KRMFTFKAF	FCILYAPSL	ILVGNLYARG	ILKQEHRLP	PGAPLAENH	
80	401	FFETVLTGSD	SHAGHLRGL	DYVSGREGWK	AKILSLDSE	LWVDEKLAD	
	451	NFLCRKYRDE	VEKAEAVFIA	QYFLRMGGQ	FVPRFAEQSF	LIPGFARFR	

501 ETVKRIAIVK PVYVFANNTS ISRSPLREEK LKRFAANQYL RPIQAMGDIG
 551 KSNQAVFDLI KDIPNVHWD AQKYLKNTV EYGRYLYGD QDHLTYFGSY
 601 YMGREFHKHE RLKSSRDGA LQ*

ORF128a and ORF128-1 show 99.5% identity in 622 aa overlap:

5	orf128a.pep	MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIIISEIQNG
	orf128-1	MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIIISEIQNG
10	orf128a.pep	SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNLYLGF
	orf128-1	SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNLYLGF
	orf128a.pep	QQGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCCCKTKTSLRVLNRNISIILFLILTA
15	orf128-1	QQGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCCCKTKTSLRVLNRNISIILFLILTA
	orf128a.pep	TSFLPSGFYTDILNQPNYYLSTLRFPPELLAGSLAVYQGTQNGRRQTANGKRLSSLC
	orf128-1	TSFLPSGFYTDILNQPNYYLSTLRFPPELLAGSLAVYQGTQNGRRQTANGKRLSSLC
20	orf128a.pep	FGALLACLFEVDKHNPFIPGMTLLPCLLTALLIRSMQYGTLPTRILSASPIVFGKISY
	orf128-1	FGALLACLFEVDKHNPFIPGMTLLPCLLTALLIRSMQYGTLPTRILSASPIVFGKISY
25	orf128a.pep	SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYLIEQPLRKRKMTFKKAF
	orf128-1	SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYLIEQPLRKRKMTFKKAF
30	orf128a.pep	FCLYLAPSLILVGYNLYARGILKQEHRLPLPGAPLAAENHFPETVLTIGDSSHAGHLRGL
	orf128-1	FCLYLAPSLILVGYNLYARGILKQEHRLPLPGAPLAAENHFPETVLTIGDSSHAGHLRGL
	orf128a.pep	DVVGSGREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFAQYDLMGGQ
35	orf128-1	DVVGSGREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFAQYDLMGGQ
	orf128a.pep	PVPRFEAQSFILPGFPAFRFETVKRIAIVKPVYVFANNTSISRSPLREEKLKRFAANQYL
	orf128-1	PVPRFEAQSFILPGFPAFRFETVKRIAIVKPVYVFANNTSISRSPLREEKLKRFAANQYL
40	orf128a.pep	RPIQAMGDIGKSNQAVFDLIKDI PNHVHWDQAQKYLKNTVEIYGRYLYGDQDHLTYFGSY
	orf128-1	RPIQAMGDIGKSNQAVFDLIKDI PNHVHWDQAQKYLKNTVEIYGRYLYGDQDHLTYFGSY
45	orf128a.pep	YMGREFHKHERLLKSSRDGALQX
	orf128-1	YMGREFHKHERLLKSSHGALQX

Homology with a predicted ORF from *N.gonorrhoeae*

50 ORF128 shows 93.4% identity over 244 aa overlap with a predicted ORF (ORF128ng) from *N. gonorrhoeae*:

	orf128.pep	VSLASVIASQIFLYEDFNQMRKTVELSAVF	30
	orf128ng	ILSEIQNGSFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTIELSTVF	112
55	orf128.pep	LSNITYLGFPQQGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCCCKTKTSLRVLNRNISI	90
	orf128ng	LSNITYLGFRGLGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCYKTKTSLRVLNRNISI	172
60	orf128.pep	ILFLILTASSFLPSGFYTDILNQPNYYLSTLRFPPELLAGSLAVYQGTQNGRRQTANGK	150
	orf128ng	ILFLILTASSFLPAGFYTDILNQPNYYLSTLRFPPELLVGSLLAVYQGTQNGRRQTENGK	232
	orf128.pep	RQLSSLCFGALLACLFEVDKHNPFIPGMTLLPCLLTALLIRSMQYGTLPTRILSASPI	210
65	orf128ng	RQLSSLCFGALLVCLFEVDKHPFIPGITLLPCLLTALLIRSMQYGTLPTRILSASPI	292

```
orf128.pep  VFVGKISYSLYLYHWIFIAFAPLRGGKQLGLPA 244
|||||
orf128na    VFVGKISYSLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYLLIEQPLRKR 352
```

5 The complete length ORF128ng nucleotide sequence <SEQ ID 833> is:

	1	ATGCAAGCTGC	TGCGATACAG	GCCCTGAATT	GACGGATTGC	GGGCGCGCGC
	51	CGTGCTATTC	GTCGATATT	TACACCTGAA	TGCGCCGAGG	CTGCGCCGAG
	101	GATTCTCTGG	GGTGACAAAT	TTCTTTGTGA	TCTCGGGATT	CCTCATATCC
10	151	AACATCATTC	TTTCTGAAAT	ACAGAACGGT	TCTTTTCTT	TCCGGGGATT
	201	TTATACCCGC	AGGATTAAAG	GGATTTATCC	TGCTTTTAT	GGCGCGGTG
	251	CCCTGGCTTC	GGGTATGTCT	TCTCAAATCT	TCTTTACGA	AGATTTCAC
	301	CAATAGAGGA	AAACCATAGA	GCTTTCTACG	TTTCTTTGT	CAATATTATTA
	351	TTTGGGGTTC	CGATTGGGGT	ATTTCGATT	GATTGCCGAG	GAGAACCCCG
15	401	TACTGCATAT	CTGGTCTTTG	GGGTAGAGG	ACAGATTAT	CCTCTGTAT
	451	CCTCTTTTTC	TGATATTCTG	TTACAAAAGA	ACCAAAATC	TACGGGTGCT
	501	GGTAAATAT	AGCATCATCC	TGTTTCTAGT	TTTGACGCGA	TCACTGTTT
	551	TGCCGCGCG	GGTTTATAAC	GACATCTCTA	ACAACCCaa	TACTATTATC
	601	CTTTGCGAC	TGAGGTTTTC	CGAGCTGTGT	TGGGGTTCG	TGTTGGCGGT
20	651	TTACGGCGCA	AGCAAAAGC	AGACAGCGA	ACACAGAAAT	GGAAACCGCG
	701	AGTTGCTTTC	TTTGGGCGAT	TTTGGGCGAT	CTTGGGCGAT	CTTGGGCGAT
	751	ATCGACAAC	ACGATTCGGT	TATCCCGGGA	AGAAACCTCG	TCCTTCCCTG
	801	CTGTCTGACG	CGCGTCTCTA	TGCGAGATAT	GTAATACGGG	ACACTTCOGA
	851	CCCGCATCTC	TGCGCGAAGC	GCATCTGTAT	TGTTCGGACA	AATCTCTTAT
25	901	TCCCTATACC	TGTACCATGT	GGATTTTATT	GCTCTCGGCC	ATTACATTAC
	951	AGGCGACAAA	CAGCTCGGAC	TGCTTCGGT	TGCGCGGTT	GCGCGGTTGA
	1001	CGCGCGGATT	TTCCCTGTGT	AGCTATTAAT	TGATTGAACA	GCCCGTTAGA
	1051	AAACGCGAAG	TGACCTTCAA	AAAGCACTTT	TTCTGCTTT	ATCTCGGCCC
	1101	GTCCTTGATG	CTTGTCTGGT	ACAACTGTGA	TTCAGAGGG	ATATTGAAAC
30	1151	AGGAACAACCT	CCGCGCCGCT	CCCGGACGCG	CGGTGTCTG	GGAAATATTC
	1201	TTTCCGGAAC	CGGTCTTGAC	CTCTCGGACG	GACACCCGCG	GACAACTGAG
	1251	GGGGTTTCTG	GATTATGTCT	GGGGCAGGGA	AGGGTGAAGA	GCTAAATATTC
	1301	TGTCCTCGTA	TTCGAGTGTG	TGGGTTTGGG	TGTGATAGAA	GCTGCGCAGC
	1351	AACCCGTTGT	GCGCAAAAAT	CGGGATGAA	GTTGAAAAAG	CCGAAGCTGT
35	1401	TTTCAATGGC	CAATTCTATG	TGTTGAGAT	GGGCGGCAG	CCCGTCCGGA
	1451	GATTTGGAAG	GCAATCTCTC	CTGATACCGG	GGTTCAAAAG	CCGATTCAAG
	1501	GAACCCGCTA	AGGAGATAGC	AGGGATCCAA	CTGTATATGT	TTTTTTCGAA
	1551	CAATACATCA	TACAGCGGTT	CTCCTTGAG	GGAGGAAATA	TTGAAAGATG
40	1601	TGCTCATCTC	TGCTCATCTC	TGCTCATCTC	TGCTCATCTC	TGCTCATCTC
	1651	AAGACAACAT	AGGCGGCTCT	TGATTTGGTT	AAAGATATTC	CGCTCTGCA
	1701	TTGGGTGGAC	GCACAAAAAT	ACCTGCCCAA	AAACACCTGT	GAATATCAAT
	1751	GACGCTATCT	TTCGCGGAC	CAGACCAACC	CGACTATTT	GGTCTTCTTG
	1801	TATATGGGGC	GGGAATTTCA	CAAACACGAA	CGCTCTCTCA	AGCATTTCCG
	1851	AGGCGGCGCA	TTCGATGAG			

This encodes a protein having amino acid sequence <SEO ID 834>:

45	1	MQAVSRYPJEI	DGLRVAVALVS	V11FHNNNRW	LPFGFLGVDI	FEVISGFLIT
	51	NIILSEVQNG	GF5RFRDFYTR	RIKRIYPAEI	AAVSLASVIA	SOIFLYEDFN
	101	QMRKTIELST	VFLSN1YLFJ	RIGLYFDLAD	ENPVLHWSL	AVEEQYLLY
	151	PLLLIFCYKK	VKSLNVLNRN	SIILELILTA	SFSLPAGFYT	LDLNPQNTY
50	201	LSTLRFPELL	VGSLVLAVGYQ	TGRNQGTEN	GKROLLSLC	FGALLVCLFV
	251	IDKHDFPIFG	ITLLPCLLCT	ALLRISMOYG	PIPTRLILAS	PIVFGVKISJ
	301	SL1LYLYLPI	AFAHYITGDK	QLGFLPAVST	AALTGAESFL	SY1LIEQNRN
	351	KRKMTFKKAF	CF1LYAPLSL	LVGNLYSRG	L1QGEHFRLL	POTVPAOIRL
	401	FPEVLGDG	SHAGHNGEL	VDVGCGREK	ACILSLDSE	LWWHDEGL
	451	NPLCRKYRDA	YFVYVQVQSA	QYDLMRGQA	LYPFLKQVSA	YFVYVQVQSA
55	501	ETVKRIAIAV	YFVYFAMNTS	TSRSPLEEK	LKRFAINQL	TR1RAMGDIG
	551	KSNQAVFDLV	KDIPNVHWD	QAQYLPKNTV	E1HGRLYLG	QDHLTYFGSY
	601	YMGREFHKHA	RLHKSRGGA	LQ*		

ORF128ng and ORF128-1 show 95.7% identity in 622 aa overlap:

60 orf128-1.pep MQAVRYRPEIDGLRAAVAVLSVMI¹FHLNNRWLPGGFLGV²DFVVISGFLITGII³LSEIQNG
orf128ng MQAVRYRPEIDGLRAAVAVLSVMI¹FHLNNRWLPGGFLGV²DFVVISGFLIT³NI⁴LSEIQNG

orf128-1.pep SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQ¹FLYEDFNQMRKTVELSAVLSNI²YLG³F
65 orf128ng SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQ¹FLYEDFNQMRKTIELSTVLSNI²YLG³F

5	orf128-1.pep	OQGYFDLSADENFVLHIWSLAVEEQYYLLYPLLIFCCCKTKSLRVLNRNISTILFLILTA :
	orf128ng	RLGYFDLSADENFVLHIWSLAVEEQYYLLYPLLIFCYKTKSLRVLNRNISTILFLILTA :
10	orf128-1.pep	SSFLPSGFTDILNQPNNTYLLSTRFPPELLAGSLAVYGOTQNGRRQTANGKRQLSSLC
	orf128ng	SSFLPAGFTDILNQPNNTYLLSTRFPPELLAGSLAVYGOTQNGRRQTANGKRQLSSLC
15	orf128-1.pep	FGALLACLFLVIDKHNPFIPGMTLLPCLLTALLIRSMQYGTLPTRILSASPPIVFGKISY
	orf128ng	FGALLVCLFLVIDKHDPFIPGITLLPCLLTALLIRSMQYGTLPTRILSASPPIVFGKISY
20	orf128-1.pep	SLYLYHWIFIAFAHYITGDKQLGLPAVSAAVATAGFSLSSYYLIEQPLRKRKMTFKKAF
	orf128ng	SLYLYHWIFIAFAHYITGDKQLGLPAVSAAVATAGFSLSSYYLIEQPLRKRKMTFKKAF
25	orf128-1.pep	FCLYLAPSLMLVGNLYARGILKQEHRLPLPGAPLAAENHFPETVLTLDGSHAGHLRGFL
	orf128ng	FCLYLAPSLMLVGNLYARGILKQEHRLPLPGTVAENHFPETVLTLDGSHAGHLRGFL
30	orf128-1.pep	DVVGSGREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
	orf128ng	DVVGSGREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
35	orf128-1.pep	PVPRFEAQSLFIPGFFARFRETVKRIAAVKPVYVFNNTSISRSPLREEKLRKFAANQYL
	orf128ng	PVPRFEAQSLFIPGFFARFRETVKRIAAVKPVYVFNNTSISRSPLREEKLRKFAINQYL
40	orf128-1.pep	RPIQAMGDIGKSNQAVFDLIKIDIPNVHWDAQYLPKNTVEIYGRYLYGDQDHLTYFGSY
	orf128ng	RFIRAMGDIGKSNQAVFDLVKIDIPNVHWDAQYLPKNTVEIYGRYLYGDQDHLTYFGSY
45	orf128-1.pep	YMGREFHKHERLLKSSHGGLQX
	orf128ng	YMGREFHKHERLLKSSHGGLQX

610 620

In addition, ORF218ng shows homology to a hypothetical *H. influenzae* protein:

40	sp P43993 Y392_HAEIN	HYPOTHETICAL PROTEIN HI0392 >gi 1074385 pir B64007 hypothetical protein HI0392 - Haemophilus influenzae (strain Rd KW20)
	>gi 1573364 (U32723)	H. influenzae predicted coding region HI0392 [Haemophilus influenzae] Length = 245
45	Score = 239 bits (604), Expect = 3e-62	
	Identities = 124/225 (55%), Positives = 152/225 (67%), Gaps = 1/225 (0%)	
50	Query: 38	VDIFFVISGFLITNIILSEIQNGSFSFRDFTYTRIKRIYFXXXXXXXXXXXXXXXXXXFLYE 97 +DIFFVISGFLIT II++EIQ SFS + FYTRIKRIY F+Y
	Sbjct: 1	MDIFFVISGFLITGIIITEIQNSFSLKQFYTRIKRIYPAFITVMALVSFIASIRIYN 60
55	Query: 98	DFNQMKTIELSTVFLSNLYGFRGLGYFDLSADENFVLHIWSLAVEEQXXXXXXXXXXIFC 157 DFN++RKTIEL+ FLN YLG GYFDLSA+ENFVLHIWSLAVE Q I
	Sbjct: 61	DPNKLARKTIELAIAFLSNFYGLTQGYFDLSANENFVLHIWSLAVEGQYYLIFPILILILA 120
60	Query: 158	YKTKTKSLRVLNRNISTILFLILTASSFLPAGFYTDILNQPNNTYLLSTRFPPELLVGSLLAV 217 YKK ++VL I++ILF IL A+SF+ A FY ++L+QPN YYLS LRFPELLVGSLLA+
	Sbjct: 121	YKFRVRVKVLFITILFLIFILLATSFVSANFYKEVLHPQNIYYLSNLRFPPELLVGSLLAI 180
	Query: 218	YGOTONGRRQTENGKRQLSSLLCGFALLVCLFLVIDKHDPFIPGIT 262 Y N + Q ++L+L L CLF+++ + FIGIT
	Sbjct: 181	YHNLN-KVQLSKQVNNILAILSTLLFSCFLMNNNIAFIGIT 224

This analysis, including the identification of several putative transmembrane domains, suggests that these proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 99

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 835>:

```

1  ..ATTATTACG AATACCGCTG GATGTTTCTT TACGGCGCAC TGACGACCTT
51  GGGCGTGACG GTCCGTGGCAA C.GCGGGCGG TTCCGATATTG GGTCTGTGTT
5  101  TGGCGTTGGC GCGCGTGATT CACTTGGAAA AACCGGTGCG GCGGATGCGC
151  GTGCTGCGCT GGGCGTTGGC TAAAGTTTCG CTGCTGTATG TTACGCTGTT
201  CCGGGGTACG CCGCTGTTTG TCAGAGATTGT GATTTGGGCG TATGTGTGGT
251  TTCCGGTTTT CGTC..

```

This corresponds to the amino acid sequence <SEQ ID 836; ORF129>:

```

10  1  ..IIYEYRMWFL YGALTTLGLT VVAXAGGSVL GLLLALARLI HLEKAGAPMR
51  VLAWALRKVS LLYVTLFRGT PLFVQIVIVA YWVFFEV..

```

Further work revealed the complete nucleotide sequence <SEQ ID 837>:

```

1  ATGGATTTC GTTTGACAT TATTACGAA TACCGCTGGA TGTTCCTTTA
51  CGGCGCACTG ACGACCTTGG GGCTGACGGT CGTGGCAACG CGGGCGGTTT
15  101  CGGTATTGGG TCTGTTGTTG GCGTTGGGCG GCTTGATTCA CTTGAAAAAA
151  GCGCGTGCGC CGATGCGCGT GCTGGCGTGG GCGTGTGGTA AAGTTTCGCT
201  GCTGTATGTT ACGCTGTTCG GGGGTACGCC GCTGTTGTG CAGATTGTGA
251  TTTGGGCGTA TGTGTGGTTT CCGTTTTCG TCCATCTCTC AGACGCGATT
301  TTGCTCAGCG GCGAGGCGCG AATCGCGCTG CGTCGCGGAT ACGGGCCGCT
20  351  GATTGCGGGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG
401  AGATTTTCCG CCGCGGCATC CAGCTCTATG ACNAAAGACA GATGGAAGCG
451  GCGCGTCTT TGGGCGTAC CTATCCGCGC GCGATGCGCT ATGTGATTCT
501  GCGCGAGGCA TTGCGCGCCA TGCTGCGCGC TTTGGGAGCG GAGTTTCATCA
551  CGCTCTTGAA AGACAGCTCG CTGCTGCGG TCATTGCTGT GGGGAGGTTG
25  601  GCGTATGTTT AGAATACGAT TACGGCGCGG TAATCGGTTT ATGAAGAACC
651  GCTTTACACC CTGCGCCCTGA TTTATCTGTT GATGACGACT TTCTAGGCT
701  GGATATTCTT GCGTTTGGAA AAACGTTACA ATCGCAACA CCGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 838; ORF129-1>:

```

30  1  MDFRFDIIYE YRMWFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
51  AGAPMRVLAW ALRKVSLLYV TLRGTPLFV QIVWAYVWF PFFVHPSDGI
101  LVSGEAAIAL RRGYGPLIAG SLALIANSGA YICETFRAGI QSIDKGQMEA
151  ARSLGLTYPQ AMRYVILFQA LRRMLPLAS EFITLLKQSS LLSVIAVAEL
201  AYVQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNFOHR*

```

Computer analysis of this amino acid sequence gave the following results:

35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF129 shows 98.9% identity over a 88aa overlap with an ORF (ORF129a) from strain A of *N.meningitidis*:

```

10 20 30 40 50
orfl29.pep IIYEYRMWFLYGALTTLGLTVVAXAGGSVLGLLLALARLIHLEKAGAPMRVLAW
40 orfl29a MDFRFDIIYEYRMWFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
10 20 30 40 50 60
60 70 80
orfl29.pep ALRKVSLLYVTFLFRGTPLFVQIVWAYVWF PFFVHPSDGI
45 orfl29a ALRKVSLLYVTFLFRGTPLFVQIVWAYVWF PFFVHPSDGI LVSGEAAIALRRGYGPLIAG
70 80 90 100 110 120
50 orfl29a SLALIANSGAYICEIFRAGIQSIDKGQMEAAARSLGLTYPQAMRYVILFQALRRMLPLAS
130 140 150 160 170 180

```

The complete length ORF129a nucleotide sequence <SEQ ID 839> is:

```

1  ATGGATTTC GTTTGACAT TATTACGAA TACCGCTGGA TGTTCCTTTA
51  CGGCGCACTG ACGACCTTGG GGCTGACGGT CGTGGGAGCG CGGGCGGTTT

```

1	01	CGGTATGTTG	TCGTGTGGTG	GGCTGGGGCG	GCCTGATTCA	TTTGGGAAGA
151	01	CGCGGTGGCG	CATGTCGGTG	GGCTGGGGTG	GGCTTGGCTA	AGGTTTCGCT
2	01	TGCGGTATGT	ACGCTGTTCG	GGGGATACGC	GGCTTTTGTG	CAGATTGTGA
251	01	TTTGGGATGT	TTGTGGTTGT	CGCTTTTGTG	TCGATTCTTG	ACAGGCGCAT
301	01	TTGGTATGCG	CGGAGCGGCG	AGATCGCGCT	CGTCGGGATG	ACGGGCGCGT
351	01	GATTCGCGGT	TCCTTGGCAC	TGATCGCCAA	CTCGGGGGCG	TATATCTGTG
401	01	AGATTTTCGC	CGCGGCATAC	CAGTCTATAG	ACAAAGGACA	GATGAGGGCG
451	03	CGCGCTTTCT	TGGGCGTCAC	CTATCGCGAC	GGGAGTCGCT	ATGCTGATTC
501	03	CGCGGAGATG	TTGCGGCGAT	CTGCTCGGCG	TCGTTGTGGT	CGGCGGATGT
551	03	CGGCGGATGC	CGGCGGATGC	TCGATGTCGT	TCGATGTCGT	CGGCGGATGT
601	03	CGCTATGTTG	AGAAATACAT	TACGCGGCGC	TTATCGGTTT	ATGAAGAACC
651	03	CGCTTACCAC	GTGCGCCTTA	TTTATCTTCT	GATGACACCT	TTCTTAGGCT
701	03	GGATATTCCT	GTGCTTGGAA	MACTGTACA	ATCGGCACTA	CCGCTTGA

15

1 MDRFDIIE YRWMFLYGLAL TTGLTVVAT AGGSVLGLLL ALARLIHLEK
51 AGAPMRVLAW ALRKVSLLYV TLFGRGTELV QIVIWAYVWF PFFVHPSDGI
101 LVSGEAAIAL RRGYCPGLAG SLALIANSGA YICEIFRAGI QSIDKGQMEA
151 ARSLGLTYPQ AMRYVILPQA LRRMLPLAS EFTITLKDSS LLSVIAVAEL
201 AVYONTITGR YSVYESEPLYT VALTYLLMTT FLGIVFLRIE KRYNPOHR

25

orf129a.pep MDPRFDIIIEYRNMFYLIGALTTLGLTVATAGGSVLGLLALARLHLEKAGAPMRVLAW
|
|
| orf129-1 MDPRFDIIEYRNMFYLIGALTTLGLTVATAGGSVLGLLALARLHLEKAGAPMRVLAW

orf129a.pep ALRKVSLLVTFLRGPTLFVOIVWAYWWFFPFVHPSDGIUVSGEAAIALRGGYGPILAG
|
|
| orf129-1 ALRKVSLLVTFLRGPTLFVOIVWAYWWFFPFVHPSDGIUVSGEAAIALRGGYGPILAG

30

orf129a.pep SLALIANSGAYICEIFRAGIQSIDKGMEARSGLGTPQAMRVYILPQALRMLPLPAS
| | | | |
orf129-1 SLALIANSGAYICEIFRAGIQSIDKGMEARSGLGTPQAMRVYILPQALRMLPLPAS
| | | | |

orf129a.pep EFITLLKDSLSLVIAVAELAVVNTTIGRYSYVEEPLYTVALIYLMTFLGWIFRLRE
| | | | |
orf129-1 EFITLLKDSLSLVIAVAELAVVNTTIGRYSYVEEPLYTVALIYLMTFLGWIFRLRE
| | | | |

35

```

orf129a.pep  KRYNPQHRX
              |||||
orf129-1     KRYNPQHRX

```

40

ORF129 shows 98.9% identity over a 88 aa overlap with a predicted ORF (ORF129ng) from *N.gonorrhoeae*:

45

orf129.pep	IIIEYRWMPFLYGALTTGLTTVVAXAGGSVLGLLLALARLIHLEKAGAPMRVLAW	54
orf129ng	MDFRFDIIIEYRWMPFLYGALTTGLTTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW	60
orf129.pep	ALRKVSLLVLTFRGTFPLFVQIWIWAYVWPFFVF	88
orf129na	ALRKVSLLVLTFRGTFPLFVQIWIWAYVWPFFVILHTAFLGNAMRQSRVDPDKGRNIAG	120

An ORF129ng nucleotide sequence <SEQ ID 841> was predicted to encode a protein having amino acid sequence <SEQ ID 842>:

55

```

1  MDFRFDIIE YRWMFLYGLA TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
51 AGAPMRVLVA ALRKVSLLYV TLFRGTPLEV QVIWAYVWF PFFVILHTAF
101 LGNMAQRSSR VPDKGRIWAG SLELNCPQPG RKTRGEFPQG ESNLGTPEPN
151 PLSMGQRRFP GCENWYPPQN FIKK*

```

Further work revealed the following gonococcal sequence <SEQ ID 843>:

1 ATGGATTTTc gttTTTGACAT TATTTA_{cga}A TACCGCTGGA TGTTTCTTTA


```

51 CGGCGCACTG Acgaccttgg gctgacgggt cgtggcgagc gCGGCGCGTT
101 CGGaatagg GCTGTGCGC GCTGATTCA CTGGAATAA
151 GCGCGTGGCG CGATGCGGCT GCTGGCGTGG CGCTTGCGTA AGGTTTCGCT
201 GCTGTACGTT ACCCTGTTCG GGGGTACGCC GCTGTTGTG CAGATTGTGA
251 TTTGGCGGTA TGTGTGGTTT CCGTTTTTCG TCCATCTCTC AGACGCGATT
301 TTGCTCAGCG GCGAGGCGCG AATCGCGCTG CGTGGCGGAT ACGGCGCGCT
351 GATTGCGCGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG
401 AGATTTTCGCG CGCGGGCATC CAGTCTATAG ACAAGAGACA GATGGAGGCG
451 GCGTGITCTT TGGGCGTACG CTATCGCGAG GCGATGCGCT ATGTGATTCT
501 GCGCGAGGCA TTGCGCGGTA TGCTGCGGCC TTTGCGGAGC GAGTTTATCA
551 CGCTCTTGAA AGACAGCTCG CTGCTGTGCG TCATTCGTCT GCGGAGTGTG
601 GCGTATGTTT AGAATACGAT TACGGGCGCG TATTGCGTTT ATGAAGAAC
651 GCTTTACACC GCGCGCGTGA TTTATCTGTT GATGACGACT TTCTTAGGCT
701 GGATATTCTT GCGTTTGGAA AAACGTTACA ATCCGCAACA CCGCTGA

```

15 This corresponds to the amino acid sequence <SEQ ID 844; ORF129ng-1>:

```

1 MDRFDIIYE YRWMFLYAL TLTGLTVVAT AGGSVLGLLL ALARLIHLEK
51 AGAPMRVLAW ALRKVSLLYV TLRGTPLFV QIVIWAYVWF FFFVHPSDGI
101 LVSGEAAIAL RRGYGPLIAG SLALIANSGA YICEIFRAGI QSIDKGMEA
151 ARSLGITYPO AMRYVILFPA LRMLPPLAS EFTLLKDDSS LLSVIAVAIL
201 AYWQVITGR YSVYEEPLNT VALIYLLMTT FLGWIFLRLE KRYNPQHR*

```

ORF129ng-1 and ORF129-1 show 99.2% identity in 248 aa overlap:

```

orfl29-1.pep MDRFDIIYEYRWMFLYALTLTGLTVVATAGGSVLGLLALARLIHLEKAGAPMRVLAW
|||||
orfl29ng-1 MDRFDIIYEYRWMFLYALTLTGLTVVATAGGSVLGLLALARLIHLEKAGAPMRVLAW
|||||
orfl29-1.pep ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFFFFVHPSDGIILVSGEAAIALRRGYGPLIAG
|||||
orfl29ng-1 ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFFFFVHPSDGIILVSGEAAIALRRGYGPLIAG
|||||
30 orfl29-1.pep SLALIANSGAYICEIFRAGIQSIDKGMEAARSLGLTYPQAMRYVILPQALRMLPPLAS
|||||
orfl29ng-1 SLALIANSGAYICEIFRAGIQSIDKGMEAARSLGLTYPQAMRYVILPQALRMLPPLAS
|||||
35 orfl29-1.pep EFTLLKDDSSLLSVIAVAILAYVQNTITGRYSVYEEPLNTVALIYLLMTTFLGWIFLRLE
|||||
orfl29ng-1 EFTLLKDDSSLLSVIAVAILAYVQNTITGRYSVYEEPLNTVALIYLLMTTFLGWIFLRLE
|||||
orfl29-1.pep KRYNPQHRX
|||||
40 orfl29ng-1 KRYNPQHRX

```

In addition, ORF129ng-1 is homologous to an ABC transporter from *A.fulgidus*:

```

2650409(AE001090) glutamine ABC transporter, permease protein (glnP)
[Archaeoglobus fulgidus]Length = 224
Score = 132 bits (329), Expect = 2e-30
Identities = 86/178 (48%), Positives = 103/178 (57%), Gaps = 18/178 (10%)

Query: 65 VSLYVTLFRGTPLFVQIVIWAYVWFFFFVHPSDGIILVSGEAAIALRRGYGPLIAGSLAL 124
+SL Y + RTPL VQI+I +F P + GI + E A G +AL
Sbjct: 58 ISTAYVEVIRGTPLLVQILI-----VYFGLPAIGINLQPEPA-----GIALL 99

Query: 125 IANSAGAYICEIFRAGIQSIDKGMEAARSLGLTYPQAMRYVILPQALRMLPPLASEFIT 184
SGAYI EI RAGI+SI GQMEAA SIG+TY QAMRYVI PQA R +LF L +EFL
Sbjct: 100 SICSGAYIAEIVRAGIESIFIGQMEAAARSLGMYLQAMRYVIFPQAFRNILPALGNEFIA 159

55 Query: 185 LLKDDSSLLSVIAVAILAYVQNTITGRYSVYEEPLNTVALIYLLMTTFLGWIFLRLEKR 242
LLKDDSSLLSVI++ EL V I P AL YL+MT L + +K+
Sbjct: 160 LLKDDSSLLSVISIVELTRVGRQIVNTTFNAWTFPLGVALFYLMITPLSRLVAYSQKK 217

```

This analysis, including the identification of transmembrane domains in the two proteins, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 100

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 845>:

```

1  . . CTGAAAGAAAT GCGCTCTGAA AGACCTGTGT TTTATTCCAA ATATCGTTTA
51  TAGAAGACATC GCCATTACTT TCCTGCTCTT GCACGCGCCG GCCGAACITTT
101 GGCTGCGCCG GCAAACGCGC GGTTTTACCG CGCTCGCGCT CGGCTTCATC
151 CTGCTCGCCA AGCTGCGTGA gCTTCACCAT CACGAACTCT TACGTAACAA
201 cTACGTCGCG ACTTATTACy TGTCTCCAACT CTTTGCCGCC GCAGgcTgAT
251 TTGTGGACAG GCGCGCGGwA ATTTCAAAC CTGCCGcYT CCGCGCCCT
301 GCACCTGATT ACCCTCGGCG GCATGATGGG CGGCGGATG ATGTTGTGGc
10 351 TGACCGCGCG ACTGTGGCAC AGCGGCTTTA CCAAACTCGA CTACCCCAAA
401 CTCTGCGCGA TTGCGCTGCC CATCCTTTTC GCGCGCGCG CTCTCGCGCG
451 TTTCTTGTgT AACGTGAACC GgTATTTTT CATTACCGTT CCGCGGATTC
501 TGACCGCGCG CGTATTGCTA CTGTATCTTT TcCGTTTAT ACCGATATTT
551 CGGCGGAATG CGTTACAGA CGATCGGAz TAr

```

15 This corresponds to the amino acid sequence <SEQ ID 846; ORF130>:

```

1  . . LKECRDKDFV FIPNIVYKNI AITFLLHAA AELWLPQTA GFALAVGEI
51  LLAKLRELHH HELLRHYVR TYLLQLFRA AGSLWTGAAX LONLPASAPL
101 HLITLGGMMG GVMVMWLTAG LWSHGFTKLD YPKLCRIAPV ILFAAAVSRA
151 FLXNVNXPFF ITVFALITAA VFVLYLEFI PIFRANAFTD DEE*

```

20 Further work revealed the complete nucleotide sequence <SEQ ID 847>:

```

1  ATGCGGCGGT TTTTCGTGCG GCGCGGGGTG CTTGCCATAC TCGGTGCGCT
51  GGTCTTTTTC ATCAACCGCG GTGCGATCGT CCTGCACCGC CAAATTTTCT
101 TGGAACTTAT GCTGCGCGCG GCATACGGCG GTTTTTTGA C TGCGGCTTTG
151 TTGGACTGGA CGGTTTTTTC GGGTAACCTG AAACCTGCG CGACTTTGAT
201 GGCGGCATTA TTGCTGCGCG CATCCGCTAT ACTGCCCTTT TCGCGCGCAA
25 251 CTGCTCTGTT TTTGCTGCGC GCCTATTGGC TGGTGTGTGT GCTGTCTGCG
301 GCGCGGCTGA TTTGCTGAGA CCGAAACACC GACAACCTCG CCCTGCTAAT
351 GTTACTTGCC GCGTTCACTG TTTTTCAGAC GGCATATGCC GTCAGCGGCG
401 ATTIGAACCT GTTGCGCGCG CAAGTGCACT TAAATATGCC GCGGCGATG
30 451 TTGCTATCCG TCGCGCTCAG TATTCTTTTG GCGCGGAAG CCTGAAAGA
501 ATGCGGCTCG AAGAACCCCTG TTTTATTCC AAATATCGT TATAAAAAA
551 TCGCAATTAC TTTCTGCTCT TTGCACGCGC CGCGCGAAT TTGGCTGCCC
601 GCGCAACCGC CGGTTTTTAC CGCGCTCGCC GTCGGCTTCA TCCTGCTGCG
65 651 CAAGCTGGGT GAGCTTACAC ATCAAGAACT CTTACGTAAA CACTACGTCC
701 GCACATTATTA CTTGCTCCAA CTCTTTGCCG CGCAGGCTTA TTTGTGGACA
75 751 GGCGCGCGGA AATTACAAA CTTGCCCGCC TCGCGCGCCC TGCACTGAT
801 TACCTCTGCG GGCATGATGG GCGGCGTAT GATGCTGCG CTGACGCGCG
851 GACTGCTGCT ACCAGCTTAT ACCAACTCG ACTACCGGAG ACTCTGCGCG
901 ATTGCGGTCX CATCTCTTTT CCGCGCGCGC GTCTCGCGG CTTTCTGAT
40 951 GAACGTGAAC CCGATATTTT TCATTACCGT TCCTGCGATT CTGACGCGCG
1001 CCGTATTGCT ACTGATCTTT TTCACGTTTA TACCGATATT TCGGGCGAAT
1051 GCGTTTACAG ACGATCOGA ATAA

```

This corresponds to the amino acid sequence <SEQ ID 848; ORF130-1>:

```

1  MRPFVGAAY LAIIGALVFF INPGAIVLHR QIFLEMLPA AYGGFILTAA
45 51  LDWTGFSGNL KPVATLMAAL LLAASAILPF SPQTASFVFA AYWLVLLEFC
101 ARLIWLDRNT DNFALLMLLA AFTVEQTAYA VSGDLNLLRA QVHLNMAAVM
151 FVSVRYSILL GAELAKECLR KDPVFIPNIV YKNIATITFL LHAAELWLP
201 AQTAGFTALA VGFILLAKLR ELIHHELLLR HYVRTYYLLQ LFAAAGYLTW
25 251 GAARKLONLPA SAPHILITLG GMMGGVMVMV LTAGLWHSFG TKLDYPKLCR
50 301 IAVPILFAAA VSRALMNVN PFIITVPAI LTAAVFVLYL FTFIPIFRAN
351 AFTDDEE*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF130 shows 94.3% identity over a 193aa overlap with an ORF (ORF130a) from strain A of *N.*

55 *meningitidis*:

[illegible]

25	1	ATGCGCGCGT	ATCTTCGTGGG	CGCGCGCGGTG	CTTGCGCATCA	TCGGTGGCGGT
	51	GGTGTTTTC	ATCAACCCCG	GTGCATCACT	GTGCACACCG	CAAAATTTCT
	101	TGGCAATTA	CGTCCGCGCG	CGATACACCG	GTTTTTCGAC	TCGGGCTTTG
	151	TTGCACTGGA	CGGGTTTTC	GATGCTGCT	AAACCTTGAT	CGACTTTGAT
	201	GGCGSCATT	TGTGTCGCG	CACTCGCAT	ACTCGCTTT	TCGCGCGAAA
30	251	CTGCTCGTT	TTTGCTGCG	GCTTATGGG	TCGTGTGCT	CGTGCTCGG
	301	CGCGCGCTGA	TTGGGTGTA	CGAAACACG	GACAACTTG	CCCTGCTAAT
	351	GTACTTGGC	GGTTCACGT	TTTTTCAGC	GAGTAATGGC	CGGGCGTAGT
	401	ATTTGAACT	GTGCGCGCG	CAAGTCATC	TAAATGATG	CGCGCGTAGT
	451	TTGCGTGG	TTGCTGCGG	TATCTTCTG	TTGCTGCGG	CGCGCGTAGT
35	501	ATGCGCTCTG	AAAGACGAG	TATTTACAT	CAATGTGCT	TATAAAGATG
	551	TGCGCATTA	CTCTGCTCT	CTGCAGCGCG	CGCGCGAAT	TTGGCTGCT
	601	CGCGAAACCG	CGGGTTTAC	CTGCTGCGC	GTGCGGCTT	TCCTGTTGGT
	651	CAAGCTCGGT	GAGCTTACC	ATCAGCAAT	CTCTGCGAAA	CATCACTGCC
	701	GCACATATTA	CTCGCTCAA	CTTTTGGCG	CGCGAGCTA	TTTGTGACAC
40	751	CGCGCGGGA	AATTACAAA	CTGCGCGCG	TCGCGGCGCT	TGCACTGAT
	801	TACCTCGGT	GCGATGATG	CTGCGGCTG	TATGCTGTG	CTGACTGCGG
	851	GACTGTGGCA	CAGCGGCTT	ACCAAGCTGT	ATCATCCGAA	ACTCTCGCGG
	901	ATCGCGCTCC	CGCTCTNTT	CGCGCGCGC	GTTCGCGGG	CTCTTTTAT
	951	GAACTGCTT	CTCAACGCT	CTCAACGCT	CGCGAGAT	CGCGAGAT
45	1001	CGTGGTGGT	GCTTTTCTG	CTGAACTTG	TACCGACTT	TCGGGCGAAC
	1051	GGSTTTACAG	ACGATACAG	ATAA		

	MRPFVGAUV	LAILGALVFA	INPGAVLVR	OIFPLETAAL	AYGGFLTAAL
50	1	LWDTFGSMAL	KPVATLPMAL	LLAASALPFL	SPOTASFVFA
	101	ARLILWDRMT	DNFALMILML	AFTVFTQTVA	VSGDNLVRLA
	151	FVSVVRSLNT	GAELAEKCRLL	KDPVFTPNVV	YKNIAITELL
	201	ATACFTQSLA	VGFIPLAKLR	EJLHHMLTCL	HYRVTRVYGL
	251	GAAKLGNLPA	SAPHLITLIG	GMGMSVMVV	LTAGLWHSFG
	301	IVAPLIFAAA	VSVRAVMNVN	PIFFITVPAI	LTAAVFLVYL
55	351	AFTDDEP*			LTVPVIFRAN

60 orf130a.pep MPFFVGAAVLALGLGVFFINPGAIVLHQRI FLEMLPAAYGGGFTLAALLDWTGFSGNL
orf130-1 MPFFVGAAVLALGLGVFFINPGAIVLHQRI FLEMLPAAYGGGFTLAALLDWTGFSGNL

orf130a.pep KPVATLMAALLLAASAILPFS POTASFFVAAYWLVL LFCARLWLDNRNTNFALLMLLA
orf130-1 KPVATLMAALLLAASAILPFS POTASFFVAAYWLVL LFCARLWLDNRNTNFALLMLLA

65 orf130a.pep AFTVPTAYAVSGDNLRLRAOVHINMAAEVFSVRVSLILGAEALKECKRLKDPVFINPVV

	orf130-1	AFTVFQYAYAVSGOLNLLRAQVHLNMAAVMFVSVRVSVILGAEALKECKRLKDPVFIPNIV	
5	orf130a.pep	YKNIAITELLHAAAEWLPAQTAGFTSLAVGFTLLAKLRELHHHELLRKHYVRTYYLLQ	
	orf130-1	YKNIAITELLHAAAEWLPAQTAGFTALAVGFTLLAKLRELHHHELLRKHYVRTYYLLQ	
	orf130a.pep	LFAAAGYLWTGAALKQNLPAASAPLHLITLGGMMGSMVMVWLTAGLWHSGETKLDYPKLCR	
10	orf130-1	LFAAAGYLWTGAALKQNLPAASAPLHLITLGGMMGSMVMVWLTAGLWHSGETKLDYPKLCR	
	orf130a.pep	IAVPILFAAAVSRVLMNVNPIFFITVPAILTAAVFLVYLLTFVPIFRANAFTDDPE	
15	orf130-1	IAVPILFAAAVSRVLMNVNPIFFITVPAILTAAVFLVYLLTFVPIFRANAFTDDPE	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF130 shows 91.7% identity over a 193 aa overlap with a predicted ORF (ORF130ng) from *N.gonorrhoeae*:

20	orf130.pep		LKECKRLKDPVFIPNIVYKNIAITELLHAA	30
	orf130ng	LNLLRAQVHLNMAAVMFVSVRVSVLLGTETLKECKRLKDPVFIPNIVYKNIAIT-LLHAA		201
	orf130.pep	AELWLPAQTAGFTALAVGFTLLAKLRELHHHELLRKHYVRTYYLLQLFAAAGSLWTGAAX		90
25	orf130ng	AELWLPAQTAGFTALAVGFTLLAKLRELHHHELLRKHYVRTYYLLQLFAAAGYLWTGAAX		261
	orf130.pep	LQNLPAASAPLHLITLGGMMGSMVMVWLTAGLWHSGETKLDYPKLCRIAVPILFAAAVSRA		150
	orf130ng	LQNLPAASAPLHLITLGGMTGSMVMVWLTAGLWHSGETKLDYPKLCRIAVSILFASAVSRA		321
30	orf130.pep	FLXNVNPEFFITVPAILTAAVFLVYLFKFIPIFRANAFTDDPE	193	
	orf130ng	VLMNVNPIFFITVPEILTAAVFMYLLTFVPIFRANAFTDDPE	364	

An ORF130ng nucleotide sequence <SEQ ID 851> was predicted to encode a protein having amino acid sequence <SEQ ID 852>:

35	1	MNKEFTHPMR	PFFVGAAVLA	ILGALVFFHQ	PRRYHPAPPN	FLGTYYAAGCI
	51	RRFFDYRFVG	PDGFFROPET	CRYFDGGVVA	CCGCFIAVET	ATCRIFRRL
	101	LAGVAAYLRL	ADLARQOHT	LRSVDVTAAF	TVFQYAYAVS	GDNLNLAQV
	151	HLNMAAVMFV	SVRVSVLLGT	ETLKECKRLD	PVFIPNIVYK	NIATILLHAA
40	201	AELWLPAQT	AGFTALAVGF	ILLAKLRELH	HHELLRKHYV	RTYYLLQLFA
	251	AAGYLWTGAA	KLQNLPAASAP	LHLITLGGMT	GGVMVMWLTA	GLWHSGETKFL
	301	DYPKLCRIAV	SILFASAVSR	AVLMNVNPIF	FITVPEILTA	AVFMYLLTFE
	351	VPIFRANAFT	DDPE*			

Further work revealed the following gonococcal DNA sequence <SEQ ID 853>:

45	1	ATGCGCCCGT	TTTTGTCGG	TGCGGCAGTA	CTTGCCATAC	TCGGTGCCTT
	51	GGTGTTTTTT	ATCAACCCCG	GGCTATCAT	CCTGCACCGC	CAAAATTTCT
	101	TGGAACITAT	GTCGCCGGCT	GCATACGGGG	GTTTTTTGAC	TACCGCTTTG
	151	TGGAACCGGA	CGGGTTTTTC	AGGCAACCTG	AAACCTGCGC	CTACTTTGAT
	201	GGCGGTGTG	TGCTTGTGTT	CGGCTGTGTT	ATTGCGGTAT	TTACCGGCAC
50	251	TTGCCGCAAT	TGTGCTGCGC	GCTCATATGC	TGCTGCTGCT	GCTGCTCTGC
	301	GCTGCTCTGA	TTGCGCTGCA	CCGCAACACC	GACACTCTCG	CTCTGTTGAT
	351	GTACTTTCGC	GCATTACCG	TTTTTCAGAC	GGCTATGCC	GTACGCGCGC
	401	ATTTGAACCT	ACTGCGCGCG	CAAGTCGATT	TGAATATGCC	GGCGCTCATG
	451	TTGCTATCCG	TCCGCGCTCAG	CGTCTTTTG	GGCACGGAAA	CCCTGAAGA
55	501	ATGCGGCTCG	AAGACCCCG	TATTCATCCC	CACGTTATTC	TATAAAACAA
	551	TGCGCATCAT	CCTGCTGCTG	CAGCGCGCGC	CCGAACTTTG	GCTGCCGCGC
	601	CAAAACCGCG	GTTTACTTGC	GCTTGCGCTC	GGCTTCATCC	TGCTGCCCAA
	651	GCTGCGCGAA	CTGCACCATC	ACGAACCTTT	ACGCAAAAC	TACGTCCGCA
	701	CTTATTACCT	GCTCCAGCTC	TTTGCGCGCG	CAGGTATTCT	GTGGCAGGCG
60	751	GGCGGCAAA	TGCAAAACCT	GCCCGCTCC	GGCGCCCTCG	ACCTGATTAC
	801	CCTCGCGCGC	ATGACGGGCT	GCGTGATGAT	GGTGTGGCTG	ACTGCCGAC
	851	TGTGCGACAG	CGGCTTTACC	AAACTCGACT	ACCCGAAACT	CTGCGCATC

```

901 GCGTCTCCA TCCTTTTCGC CTCGCGCGT TCGCGCGGT TTTTAATGAA
951 CGTGATCCG ATATTCTCA TCACGCTCC CGAGATTCTG ACCGCGCGCG
1001 TGTTTCATGCT TTACCTGCTG ACGTTGCTGAC CGATTTCCTG AGCGAACGCG
1051 TTTACAGACG ATCCGAATA A

```

5 This corresponds to the amino acid sequence <SEQ ID 854; ORF130ng-1>:

```

1 MRPFVGAIV LAILGALVFF INPGAII LHR QIFLEIMLPA AYGGFLTTLAL
51 LDRTGFGSGL KPAATLMMAVL LLVAALLPF LPOLAAFFVA AYWLVLVLLFC
101 AWWIWLDRNT DNFAILLMLLA AFTVFQTAAYA VSGDNLILRA OVHNLMAAVM
151 FVSVRVSVLL GTETLKECLR KDPVFIPNVI YKNIATILL HAAAEIWLPA
201 QTAGFTALAV GILLAKLRE LHHHELLRKH YVRTYLLQL FAAAGYIWTG
251 AAKLQNLPA PASPLHLITLGG MTGGVMVWL TAGIWHSGFT KLDYPKLCIRI
301 AVSILFASAV SRAVLMNVNP IFFITVPEIL TAAVFMLYLL TFVPIFRANA
351 FTDDPE*

```

ORF130ng-1 and ORF130-1 show 92.4% identity in 357 aa overlap:

```

15 orf130-1.pep MRPFVGAIVLAILGALVFFINPGAIVLHRQIFLEIMLPAAYGGFLTALLDWTGFGSGL
orf130ng-1 MRPFVGAIVLAILGALVFFINPGAII LHRQIFLEIMLPAAYGGFLTALLDRTGFGSGL
20 orf130-1.pep KPVATLMAALLLAASAILPFSPTASFFVAAYVLVLLFCALINLDRNTDNFALLMLLA
orf130ng-1 KPAATLMVALLVAAVLLPFLPOLAAFFVAAYWLVLVLLFCANLWLDRNTDNFALLMLLA
25 orf130-1.pep AFTVFQTAAYVSGDNLILRAQVHNLMAAVMFVSVRVSLILGAEALKECLRKDPVFIPNIV
orf130ng-1 AFTVFQTAAYVSGDNLILRAQVHNLMAAVMFVSVRVSLILGTECLKELRKDPVFIPNVI
30 orf130-1.pep YKNIATIFLLHAAAEIWLPAQTAGFTALAVGFILLAKLRELHHELLRKHVYRTYLLQL
orf130ng-1 YKNIAT-ILLHAAAEIWLPAQTAGFTALAVGFILLAKLRELHHELLRKHVYRTYLLQL
35 orf130-1.pep LFAAGYIWTGAARKLQNLPAASPLHLITLGGMTGGVMVWLTAGIWHSGFTKLDYPKLCIR
orf130ng-1 LFAAGYIWTGAARKLQNLPAASPLHLITLGGMTGGVMVWLTAGIWHSGFTKLDYPKLCIR
40 orf130-1.pep IAVPIILFAAAVSRAEIMNVNPIFFITVPEILTAFAVFLYLLTFVPIFRANAFTDDPEX
orf130ng-1 IAVSILFASAVSRAVLMNVNPIFFITVPEILTAFAVFLYLLTFVPIFRANAFTDDPEX

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

40 Example 101

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 855>:

```

1 ATGGAATTC GGGCAATAA ATATACGGCA ATGGCTGCGT TGCTTGCATT
51 TACGGTTGCA GGCTGCGGCC TGGCGGGGTG GTATGAGTGT TCGTCCCTCA
101 CCGGCTGGTG TAAGCCGAGA AAACCGCGCT CCATCGATT TGGGATATT
151 GCGCGCGAGA GTCGCGCCCT TTTAGGGGAC TACGAGATAC CGCTTTCAGA
201 CGGCAATAGT TCCGTCAGGG CAACGAATA TGAATCCGCA CACAATCTT
251 ACTTTTACAG GAAATATAGG AAGTTTGAAG C.TGCGGGCT GGATTGGCGT
301 ACGCGTGACG CAAACCTTT GATTGAGAGC TTCACACAGG GAGGATTGTA
351 CTGCTTGGA AAG.

```

50 This corresponds to the amino acid sequence <SEQ ID 856; ORF131>:

```

1 MEIRAIKYTA MAALLAFTVA GRLAGWYEC SSITGWCKPR KPAADFWDI
51 GGESPPSLGD YEIPLSDGNS SVRANEYESA QSSYFYRKIG KFEXCGLDWR
101 TRDGKPLIET FKQGGFDCL K.

```

Further work revealed the complete nucleotide sequence <SEQ ID 857>:

```

55 1 ATGGAATTC GGGCAATAA ATATACGGCA ATGGCTGCGT TGCTTGCATT
51 TACGGTTGCA GGCTGCGGCC TGGCGGGGTG GTATGAGTGT TCGTCCCTCA

```

101 CCGGCTGGTG TAAGCCGAGA AACCCTGGTG CCATCGATT TTGGGATATT
 151 GGGCGCGAGA GTCCGCGTC TTAGGGGAC TACGAGATAC CGCTTTCAGA
 201 CGGCAATCGT TCCGTCAGG CAAACGAATA TGAATCCGCA CAACAATCTT
 251 ACTTTTACAG GAAATAGGG AAGTTTGAAG CCTGCGGGCT GGATTGGCGT
 301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAACACAGG GAGGATTTGA
 351 CTGCTTGGAA AAGCAGGGGT TCGGCGGCAA CGGCTGTGCC GAGCGCGTCC
 401 GATGGTAA

This corresponds to the amino acid sequence <SEQ ID 858; ORF131-1>:

1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLTGWCKPR KPAADFWDI
 51 GGESPPSLGD YEIPLSDGNR SVRANEYESA QOSYFYRKIG KFEACGLDWR
 101 TRDGKPLIET FKQGGFDCL KQGLRRNGLS ERVRW*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF131 shows 95.0% identity over a 121aa overlap with an ORF (ORF131a) from strain A of *N.*

15 *meningitidis*:

		10	20	30	40	50	60
orf131.pep	MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAADFWDIGGESPPSLGD						
orf131a	MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAADFWDIGGESPPSLGD						
		10	20	30	40	50	60
		70	80	90	100	110	120
orf131.pep	YEIPLSDGNSSVRANEYESAQOSYFYRKIGKFEXCGLDWRTRDGKPLIETFKQGGFDCL						
orf131a	YEIPLSDGNSSVRANEYESAQOSYFYRKIGKFEXCGLDWRTRDGKPLIETFKQGGFDCL						
		70	80	90	100	110	120
orf131.pep	K						
orf131a	KQGLRRNGLSERVRW						
	130						

The complete length ORF131a nucleotide sequence <SEQ ID 859> is:

1 ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCATT
 51 TACGGTTGCA GGCTGCCGGT TGGCAGSTTG GTATGAGTGT TCGTCCCTGT
 101 CCGGCTGGTG TAAGCCGAGA AACCCTGGTG CCATCGATT TTGGGATATT
 151 GGGCGCGAGA GTCCCTCCGT TTAGAGGAC TACGAGATAC CGCTTTCAGA
 201 CGGCAATCGT TCCGTCAGG CAAACGAATA TGAATCCGCA CAACAATCTT
 251 ACTTTTACAG GAAATAGGG AAGTTTGAAG CCTGCGGGTT GGATTGGCGT
 301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAACACAGG AAGGTTTGA
 351 TTGTTGAAA AAGCAGGGGT TCGGCGGCAA CGGCTGTGCC GAGCGCGTCC
 401 GATGGTAA

This encodes a protein having amino acid sequence <SEQ ID 860>:

1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLTGWCKPR KPAADFWDI
 51 GGESPPSLGD YEIPLSDGNR SVRANEYESA QOSYFYRKIG KFEACGLDWR
 101 TRDGKPLIET FKQGGFDCL KQGLRRNGLS ERVRW*

ORF131a and ORF131-1 show 97.0% identity in 135 aa overlap:

orf131a.pep	MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAADFWDIGGESPPSLGD
orf131-1	MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAADFWDIGGESPPSLGD
orf131a.pep	YEIPLSDGNSSVRANEYESAQOSYFYRKIGKFEXCGLDWRTRDGKPLIETFKQGGFDCL
orf131-1	YEIPLSDGNSSVRANEYESAQOSYFYRKIGKFEXCGLDWRTRDGKPLIETFKQGGFDCL
orf131a.pep	KQGLRRNGLSERVRW

orf131-1 KCGLRRNGLSERVRWX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF131 shows 89.3% identity over 121 aa overlap with a predicted ORF (ORF131ng) from

5 *N.gonorrhoeae*:

	orf13l.pep	MEIRAKYTTAMAALLASTVACRLAGWVECSLSLIGCKFKRPAATDFWFIDGESPSLG	60
	orf13l.ng	MEIRVIKYYTATALEAFVACRLAGWVECLSLSGWKFKRPAATDFWFIDGESPSLED	60
10	orf13l.pep	YEIPLSDGNSSVRANEYESAQQSYFYRKIGKFECGLDWRTRDGKLIETFKQGGFDCLE	120
	orf13l.ng	YEIPLSDGNSSVRANEYESAQKSIFYRKIGKFECGLDWRTRDGKLIVERFKQGGFDCLE	120
15	orf13l.pep	K KOGLRRNGLSERVRW	121 134
	orf13l.ng		

A complete length ORF131ng nucleotide sequence <SEQ ID 861> was predicted to encode a protein having amino acid sequence <SEO ID 862>:

20 1 MEIRVIKYTA TAALFAFTVA GCRLAGWYEC LSLSGWCKFR KPAAIDFWDI
51 GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYFYRKIG KFEACGLDWR
101 TRDGKPLVER FKOEGFDCLE KOGLRRNGLS ERVRW*

Further work revealed the following gonococcal DNA sequence <SEQ ID 863>:

		ATGGAATTC	GGGTATAAAA	ATATACGGCA	ACGGCTCGCT	TGTTTGCATT
25	51	TACGCTTGCA	GGCTCCGGC	TGGCGGGCTG	GTA TAGGT	TGCTCTTGT
	101	CCGGCTGGTG	TAA GCGAGA	ATACCTGCG	CCATCGATT	TGGGATGATT
	151	GCGCGAGAGA	TGcgcctGT	TTTAGAGAGC	TACAGATATC	CGCTTCACGA
	201	CGCAATCTGT	TCCTG CAGG	CACAAGAATAT	TGAATCGCTG	CAAAATCTCT
	251	ACTTTTATAG	GA AAAATAG	AAGT TTGAAG	CTCTGGGGCT	GAGATGGCT
	301	ACCGCTGACG	CGAAACCTTG	GTTTAGAGAG	TTCAACAGC	AGAGTTTTCG
30	351	CTTTTGGAA	ANGCA GGGT	TGCGGCGCAA	CGGCTCTGTC	GACGCGCTCA
	401	GATGGTAA				

This corresponds to the amino acid sequence <SEQ ID 864; ORF131ng-1>:

35 1 MEIRVIKYTA TAALFAFTVA GCRLAGWYEC SSLSGWCKPR KPAAIDFWDI
51 GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYFYRKIG KFEACGLDWR
101 TRDGKPLVER FKOEGEDCLE KOGLRRLGLS ERVRW*

ORF131ng-1 and ORF131-1 show 92.6% identity in 135 aa overlap:

orf131ng-1.pep MEIRVIKYTTATAALFATVAGCRLAGWECSSLSGWCKPRKPAADFDWDIGGESPLSLSD
| | | | |
orf131-1 MEIRAIKYTAMAAALFATVAGCRLAGWECSSLTGWCKPRKPAADFDWDIGGESPFSLGD
| | | | |

orf131ng-1.pep YEIPLSDGNRSVRANEYESAQSKSYFFRKIGKFEACGLDWRTRDGKPLVERFKQEGDFCL
| | | | |
orf131-1 YEIPLSDGNRSVRANEYESAQSQSYFFRKIGKFEACGLDWRTRDGKPLIEFTKGGGDFCL
| | | | |

orf131ng-1.pep KQGLRRNGLSERVRWX
| | | | |
orf131-1 KQGLRRNGLSERVRWX
| | | | |

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 102

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 865>

```

1 ATGAAACACA TCCATATTAT CGGTATCGGC GGCACGTTTA TGGCGGGGCT
5 51 TGCGGCCATT GCCAAGAAG CGGGGTTTGA AGTCAGCGGT TCGCAGCGGA
101 AGATGTATCC GCGCATGAGC ACCCAGCTCG AAGCCTTGGG TATAGACGTG
151 TATGAAGGCT TCGATGCCGC TCAGTTGGAC GAATTTAAG CCGAGCTTTA
201 CGTTATCGGC AATGTGCGCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT
251 TGAACCTCGG CTGCGCTTAT ATTtCGGCC CGCAATGGCT GTCGGAACAC
301 GTGCTGCACC ATCATTGGGT ACTCGGTGTG GCGGGAGACG ACGGCAAAAC
10 351 GACCACCGCC TCCATGCTCG CATGGGTCTT GGAATATGCC GGCCTCGCGC
401 CGGGCTTCCT TATtGGCGGC GTACC . GGAA AATtCGGCG TTTCGCGCCG
451 CCTGCCGCAA ACGCCCGGCC AAGACCCGAA CAGCCAAATCG CCGTTTTTCG
501 TCATCGAAGC CGACGAATAC GACACCGCCT TTtCGACAA ACCTTCTAAA
15 551 TtCGTCATT ACCGTCCGCG TACCGCCGTG TTGAACAATC TGAATTCGA
601 CCACGCCGAC ATCTTTGGCG ACTTGGCGCG GATACAGACC CAGTTCACCT
651 ACCTCGTGGC TACCGTGGCG TCTGAAGGCT TAATCGTCTG CAACGGACGG
701 CAGCAAGGCC TCGAAGATAC TTTGGACAAA GGCTGCTGGA CGCCGGTGGG
751 AAAATTCGGC ACGGAACACG GCTGCA . .

```

This corresponds to the amino acid sequence <SEQ ID 866, ORF132>:

```

20 1 MKHIHIIGIG GTFMGGGLAAI AKEAGFEVSG CDKMYPPFMS TQLEALGIVD
51 YEGFDAQLD EFKADVYVIG NVAKRGMDVV EAILNLGLPY ISGPQWLSEN
101 VLHHHVVVLV AGTHGKTTTA SMLAWVLEYA GLAPGLLIG VVGKRRFRFP
151 PAANAAAPPE QPIAIVRHRS RRIHRLFRQ TFXIRALPSA YRRVEQSGIR
201 FRRHLRLRGR DTFVFLPRA YRAVXRLNRL QRTAAKPAFY FGRLLDAGG
25 251 KIRHGRRLA . .

```

Further work revealed the complete nucleotide sequence <SEQ ID 867>:

```

1 ATGAAACACA TCCATATTAT CGGTATCGGC GGCACGTTTA TGGCGGGGCT
5 51 TGCGGCCATT GCCAAGAAG CGGGGTTTGA AGTCAGCGGT TCGCAGCGGA
101 AGATGTATCC GCGCATGAGC ACCCAGCTCG AAGCCTTGGG TATAGACGTG
151 TATGAAGGCT TCGATGCCGC TCAGTTGGAC GAATTTAAG CCGAGCTTTA
201 CGTTATCGGC AATGTGCGCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT
251 TGAACCTCGG CTGCGCTTAT ATTtCGGCC CGCAATGGCT GTCGGAACAC
301 GTGCTGCACC ATCATTGGGT ACTCGGTGTG GCGGGAGACG ACGGCAAAAC
35 351 GACCACCGCC TCCATGCTCG CATGGGTCTT GGAATATGCC GGCCTCGCGC
401 CGGGCTTCCT TATTGGCGGC GTACCAGCAA ATTTCGCGGT TTCCGCCCGC
451 CTGCGCGCAA CGCCGCGCC AGACCCGAA AGCAATCGC CTTTTTCGT
501 TCATCGAAGCC GACGAATACG ACACCGCCTT TTTCGACAAA CCGTCAAAAT
551 TCGTGCAATTA CCGTCCGCGT ACCCGCGGTG TGAACAATCT GGAATTCGAC
601 CACGCGGACA TCTTTTGGCA CTGGGCGCG ATACAGACC AGTTCACATA
40 651 CCTCGTGGCT ACCGTGCCGT CTGAAGGCTT AATCGTCTG AACGACGCGC
701 AGCAAGCCT CAGAAGATACT TTGGACAAGG GCTGCTGGAC GCCGCTGGAA
751 AAATTCGGCA CGGAACACGG CTGGCAGGCC GCGCAAGCCA ATGCCGACGG
801 CTCGTTCGAC GTGTGCTCG ACGGCAAAAC CGCCGGACGC GTCAAATGGG
851 ATTTGATGGG CAGGCACAAAC CGCATGAACG CGCTCGCGGT CATTGCCGCC
45 901 GCGGCTCATG TCGGTGTCGA TATTAGACC GCCTGCGAAG CTTGGGCGC
951 GTTTAAAAAC GTCAAACGCC GGATGGAAT CAAAGGACG GCAACCGCA
1001 TCACCGTTTA CGACGACTTC GCCCACCACC GCACCGCAT CGAAACACG
1051 ATTCAGGTTT TGCGCCAAGC CGTGGCGGCG GCGCGCATCC TCGCCCTCTC
1101 GCAACCGCGT TCCAACACGA TGAAGCTGGG CAGATGAAG TCGCCCTGCG
50 1151 CTGTGAAGCTT CAAAGAAGCC GACCAAGTGT TCTGCTACG CGCGCGCGTG
1201 GACTGGGACG TCGCCGAAGC CTCGCGCCT TTGGCGCGGA GCGTGAACGT
1251 CGCAAGGAC TCGATGCTCT TCGTTGCCA AATCGTGAAA AACGCCAAG
1301 TAGCGACCA TATTTTGTGT ATGACCAACG CGGTTTCGCG CGGAATACAC
1351 GGAAGCTGCG TGAAGCTTT GAGATAG

```

55 This corresponds to the amino acid sequence <SEQ ID 868, ORF132-1>:

```

1 MKHIHIIGIG GTFMGGGLAAI AKEAGFEVSG CDKMYPPFMS TQLEALGIVD
51 YEGFDAQLD EFKADVYVIG NVAKRGMDVV EAILNLGLPY ISGPQWLSEN
101 VLHHHVVVLV AGTHGKTTTA SMLAWVLEYA GLAPGLLIG VPENFGVSR
151 LPOTPRODPN SOSFFVIEA DEYDTAFDFK RSKFVHYRFR TAVLNLEFD
60 1601 HADIFADLGA IQTFHYLVR TVPSEGLIVC NGRQSLQDT LDKGWTFPE
251 KFGTEHWQA GEANADSGFD VLLDGKTAGR VKWDLMRHNR RNALAVIAA
301 ARHVGVDIQT ACEALGAFKN VKRMEIKGT ANGITVYDDF AHHPATAETT

```



```

351 IQGLRQRVGG ARILAVLEPR SNTMKLGTMK SALPVSLSKEA DQVFCYAGGV
401 DWDVAEALAP LGRRLNVGKD FDAFAETIVK NAEVGDHILV MSNGGFGGIH
451 GKLEALR*

```

Computer analysis of this amino acid sequence gave the following results:

5 Homology with the hypothetical o457 protein of *E.coli* (accession number U14003)

ORF132 and o457 show 58% aa identity in 140 aa overlap:

```

10 Orf132: 4  IHIIGIGGTFMGGGLAAIAKEAGFEVSGCDKMYPPMSTQLEALGIDVYEGFDAQAQDEFK 63
    IRI+GI GTFMGGGLA +A++ G EV+G DA +YPPMST LE GI++ +G+DA+QL+ +
o457: 3  IHIILGIGTFMGGGLAMLARQLGHEVTGSDANVYPPMSTLLEKQGLIELIGYDASQLEP-Q 61

Orf132: 64  ADVVYIGNVAKRGMDVVEAILNLGLPYSIGPQWLSENVLHHHWLVGAGTHGKTTTASML 123
    D+ +IGN RG VEA+L +PY+SGPQWL + VL VWL VAGTHGKTTTA M
o457: 62  PDLVIIGNAMTRGNPCVEAVLEKNIPYMSGPQWLHDFVLDRWVLA VAGTHGKTTTATGMA 121

15 Orf132: 124 AWWLEYAGLAPGLFVIGVXG 143
    W+LE G PGF+IGGV G
o457: 122 TWILEQCGYKPGFVIGGVPG 141

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

20 ORF132 shows 74.6% identity over a 189aa overlap with an ORF (ORF132a) from strain A of *N.*

meningitidis:

```

25 orf132.pep  MKHIIHIIIGIGTFMGGGLAAIAKEAGFEVSGCDKMYPPMSTQLEALGIDVYEGFDAQAQD
    orf132a     MKHIIHIIIGIGTFMGGGLAAIAKEAGFEVSGCDKMYPPMSTQLEALGIDVYEGFDAQAQD
                10      20      30      40      50      60
                10      20      30      40      50      60

30 orf132.pep  EFKADVYIGNVAKRGMDVVEAILNLGLPYSIGPQWLSENVLHHHWLVGAGTHGKTTTA
    orf132a     EFKADVYIGNVAKRGMDVVEAILNLGLPYSIGPQWLSENVLHHHWLVGAGTHGKTTTA
                70      80      90      100     110     120
                70      80      90      100     110     120

35 orf132.pep  SMLAWVLEYAGLAPGLFVIGVXGKFR---RFRPPAANAAPFEQPI-----AVFR
    orf132a     SMLAWVLEYAGLAPGLFVIGVXGKFR---RFRPPAANAAPFEQPI-----AVFR
                130     140     150     160
                130     140     150     160

40 orf132.pep  HRSRRIRHRLFRQTFXIRALPSAYRRVEQSGIRPRRLCLGRDTPVPLFRAYRAVXRL
    orf132a     KRKSFVHYRPRPTAVLNNLEFDHADIFADLGAIQTFPHLVRTVPSEGLIVCNGRQSLQD
                170     180     190     200     210     220
                180     190     200     210     220     230

```

45 The complete length ORF132a nucleotide sequence <SEQ ID 869> is:

```

1  ATGAACACACA  TCCACATTAT  CGGTATCGGC  GGCACGTTTA  TGGGTGGGAT
51  TGCCGCCAATT  GCCAAGAAG  CAGGGTTTGA  ANTACAGCGT  TGCAGTCGCA
101 AGATGTATACC  GCGCATGAGC  ACCCAGCTCG  AAGCCTTGGG  CATAGCGCGT
151 TATGAAGGCT  TCGACACCGC  CGAGTTGGAC  GAATTTAAG  CCGACGTTTA
201 CGTTATCGGC  AATGTGCGCA  AGCGCGGGAT  GGATGTGGTT  GAAGCGATTT
251 TGAACCGTGG  GCTGCTTAT  ATTTCCGGCC  CGCAATGGCT  GGCTGAAAC
301 NTGCTGCACC  ATCAITGGNN  ACTCGSGGTG  CGCGNGACGC  ACGGCAAAAC
351 GACCACCGCG  TCTATGCTCG  CGTGGGTTT  GGAATATGCC  GGACTCGCAC
401 CGGCTTCNT  GTACCGCGGC  GTACCGGAAC  ACTTCAGCGT  TCCGCGCCG
451 TCGCCGCAAA  CGCCGCGGCA  AGACCGGACA  AGCCATGCG  CCTTTTGGT
501 CATTTGAAGCC  GACGARTACG  ACACCGCGTT  TTTGACAAAC  CGCTCCCAAT
551 TCGTGCAATTA  CCGTCCGCGT  ACCGCGGTG  TGAACAATCT  GGAATTCGAC
601 CACGCGGACA  TCTTCCGCGA  TTTGGGCGCG  ATACAGACCC  AGTTCCACCA
651 CCTCGTGGT  ACCGTGCGGT  CTGAAGGCGT  CATCGTCTGC  AACGACGCGC
701 AGCAAAAGCT  GCAAGCACT  TTGGACAAAG  GCTGCTGGAC  CGCCGTGGAA
751 AAATTCGGCA  CGGAACACGG  CTGGCAGGCC  GCGGAGGCCA  ATCCCGATGG

```

-470-

801 CTGGTTCGAC GTGTTGCTTG ACGCCAAAA AGCCGGACAC GTCCGTTGGA
 851 GTTTGATGGG CGGACACAAC CGCATGAACG CGCTCGNGT CATCGCCGCC
 901 GCGCGTTCAT CGGAGTNGA CATTGACAGC GCCTGCGAAG CCTTGAGCAC
 951 GTTTAAAAAC GTCAACAGCC GCATGGAAT CAAAGCCACG GCAACGCTA
 1001 TCACCGTTTA CGACGACTTC GCCCACCATC CGACCGCTAT GCAACACACG
 1051 ATTCAAGGTT TGGCCAGCG CGTCGGCGCG CGCGCGTCTT TCGCGTCTCT
 1101 CGAACCGCGT TCAATACGA TGAAGCTGGG TACGATGAAA GCCGCCCTGC
 1151 CCGCAAGCGT CAAAGAAGCC GACCAAGTGT TCTGNTACGC CGCGCGCGCG
 1201 GACTGGGAGC TTGCCGAAGC CCTCGCGCCT TTGGCGGCGA GGCTGCACGT
 1251 CGGCAAGAC TTGATGCGCT TCGTTGCGA AATCGTGAAA AACGCGGAG
 1301 CAGCGACCA TATTTTGGTG ATGAGCAACG GCGGTTTCGG CGGAATACAC
 1351 ACCAACTGC TGGACGCTT GAGATAG

This encodes a protein having amino acid sequence <SEQ ID 870>:

1 MKHIHIIGTG GTFMGGIAAI AKEAGFEXSG CDARKYPPMS TQLEALGIVG
 51 YEGFDTAQLD EFKADVYVIG NVAKRGMDDV EAILNRGLFY ISGPQWLAEN
 101 XLHHWKLGV AXTHGKTTTA SMLAWLEYA GLAPGFXIGG VENFVSAR
 151 LPQTFRQDFN SQSPFFVIEA DEYDTAFDD RSKFVHYRER TAVLNLEFD
 201 HADIADLGA IOTQFHLVR TVSEGLIVC NSRQSLQDT LDKGCTPVE
 251 KFGTEHGWQA GEANADGSD VLLDGKAGH VAWSLMGHIN RNNALAVIAA
 301 ARHAGVDIQT ACEALSTFFKN VKRMEIKGT ANGTVYDDF AHHPAIAETT
 351 IQGLRQRVGG ARILAVLEPR SNTMKLGTAK AALPASLKEA DQVFXAGGA
 401 DWDVAEALAP LGGRHLVVKD FDFVAVIEIK NAEAGDHILV MNGGFGGHIH
 451 TKLLDALR*

ORF132a and ORF132-1 show 93.9% identity in 458 aa overlap:

25 orf132a.pep MKHIHIIGTGTFMGGIAAIAKEAGFEXSGCDARKYPPMSTQLEALGIVYEGFDTAQLD
 orf132-1 MKHIHIIGTGTFMGGIAAIAKEAGFEVSGCDARKYPPMSTQLEALGIDVYEGFDAAQLD
 30 orf132a.pep EFKADVYVIGNVAKRGMDDVEAILNRGLFYISGPQWLAENXLHHWKLGVAXTHGKTTTA
 orf132-1 EFKADVYVIGNVAKRGMDDVEAILNLGLPYISGPQWLESENVLHHHWLGVAGTHGKTTTA
 orf132a.pep SMLAWLEYAGLAPGFXIGGVPEMFVSARLPQTFRQDFNSQSPPFVIEADEYDTAFDDK
 35 orf132-1 SMLAWLEYAGLAPGLIGGVPEMFVSARLPQTFRQDFNSQSPPFVIEADEYDTAFDDK
 orf132a.pep RSKFVHYRERTAVLNLEFDHADIADLGAIQTFHHLVRTVPSEGLIVCNRQSLQDT
 40 orf132-1 RSKFVHYRERTAVLNLEFDHADIADLGAIQTFHHLVRTVPSEGLIVCNRQSLQDT
 orf132a.pep LDKGCTPVEKFGTEHGWQAGEANADGSDVLLDGKAGHVAWSLMGGHNRNNALAVIAA
 orf132-1 LDKGCTPVEKFGTEHGWQAGEANADGSDVLLDGKTAGRVKNDLMGRHNRNNALAVIAA
 45 orf132a.pep ARHAGVDIQTACEALSTFFKNVKRMEIKGTANGITVYDDFAHHPAIAETTIIQGLRQRVGG
 orf132-1 ARHAGVDIQTACEALGAFKNVKRMEIKGTANGITVYDDFAHHPAIAETTIIQGLRQRVGG
 50 orf132a.pep ARILAVLEPRSNTMKLGTAKAALPASLKEADQVFXAGGADWDVAEALAPLGGRHLVVKD
 orf132-1 ARILAVLEPRSNTMKLGTAKAALPASLKEADQVFXAGGADWDVAEALAPLGGRHLVVKD
 orf132a.pep FDFVAVIEKNAEAGDHILVMSNGGFGGIHTKLLDALRX
 55 orf132-1 FDFVAVIEKNAEAGDHILVMSNGGFGGIHTKLLDALRX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF132 shows 89.6% identity over 259 aa overlap with a predicted ORF (ORF132ng) from *N. gonorrhoeae*:

60 orf132.pep MKHIHIIGTGTFMGGIAAIAKEAGFEVSGCDARKYPPMSTQLEALGIDVYEGFDAAQLD 60
 orf132ng MKHIHIIGTGTFMGGIAAIAKEAGFEVSGCDARKYPPMSTQLEALGIVHEGFDAAQLE 60

-471-

5	orf132.pep	EFKADVVYVIGNVAKRGMDVVEAILNLGLPYISGPWLSENVLHHHVLGVAGTHGKTTTA	120
	orf132.ng	EFQADIYVIGNVARRGMDVVEAILNRGLPYISGPWLAENVLHHHVLGVAGTHGKTTTA	120
10	orf132.pep	SMIAWVLEYAGLAPGFLIGGVXGKFRFRFPANAAPRPEQPIAVFRHRSRRIRHRLFRQ	180
	orf132.ng	SMIAWVLEYAGLAPGFLIGGVXGKFRFRFPANAASRPEQPIAVFRHRSRRIRHRLFRQ	180
15	orf132.pep	TFXIRALPSAYRRVEQSGIRPRRLHRLGRDTPVPLPRAYRAVXRNLRLQRTAAKPARY	240
	orf132.ng	TLQIRALSPAYRRVEQSGIRPRRLHRLGRDTPVPPRAHRTIRRPRLQRTAAKPARY	240
20	orf132.pep	FGORLLDAGGKIRHGRTRLA	259
	orf132.ng	FGORLLDAGGKIRHRLRLADW	261

An ORF132ng nucleotide sequence <SEQ ID 871> was predicted to encode a protein having amino acid sequence <SEQ ID 872>:

20	1	MKHHIIIGIG	GTFMGGIAAI	AKEAGFKVSG	CDAKMYPPMS	TQLEALGIGV
	51	HEGFDAQAQLE	EFQADIYVIG	NVARRGMDVV	EAILNRGLPY	ISGFWLAENV
	101	VLHHHVLGV	AGTHGKTTTA	SMIAWVLEYA	GLAPGFLIGG	VFGKFRFRFP
	151	FPANAASRPE	QPIAVFRHRS	RRIRHRLFRQ	TLQIRALSPA	YRRVEQSGIR
	201	PRRLHRLGR	DTDFVPPFRA	HRTIRRPRL	QRTAAKPARY	FGORLLDAGG
	251	KIRHRLRLAD	#*			

Further work revealed the following gonococcal DNA sequence <SEQ ID 873>:

25	1	ATGAACACACA	TCCACATTAT	CGGTATCGGC	GGCAGCTTTA	TGGGCGGGAT
	51	TGCCGCCATT	GCCAAAGRAAG	CCGGGTTCAA	AGTCAGCGGT	TGCAGCGGTA
	101	AGATGTATCC	GCGGATGAGC	ACCCAGCTCG	AAGCCTTGGG	CATAGGCGTA
	151	CACGAAGCTT	TGATGCGCG	GCAGTTGGAA	GAATTCAAG	CCGATATTTA
	201	CGTACGCGGC	AATGTCGCCA	GGCGCGGAT	GGATGTGGT	GAGGCGATTT
	251	TGAACCGTGG	GTCGCTTAT	ATTTCGGCC	CGCAATGGCT	GGCTGAAAC
30	301	GTGCTgcacc	atcaTTGGgt	ACTGCGCGTG	GcagggaCGC	ACGGcaaaAC
	351	gaccaCcGcg	tCCATGCTCG	CTGGGTCTT	GGAATATGCC	GGACTCGCGC
	401	CGGCGTCTCT	CATCGCGGgt	gtaccggaAA	ATTTCGGGgt	TTCCGCGCGC
	451	CTACCGCAAA	CGCCGCGTCA	AGACCGGAAC	AGCAAAATGG	CGTTTTTGGT
	501	CATCGAAGCC	GACGAATACG	ACACCGCTTT	TTTCGACAAA	CGCTCCAAAT
	551	TCGTGCGATT	TGCGCCCGGT	ACCGCGGTGT	TGAACAATCT	GGAAATTGAC
35	601	CACGCGGACA	TCTTCGCGCA	CTTGCGGCGG	ATACAGACCC	AGTTCGACCA
	651	CTCGTGTGCG	ACGCTACCAT	CCGAGGCGCT	CATGCTCTCG	AACGCAACGC
	701	AGCAAAAGCT	GCAAGATAC	TTGGACAAG	GCTGTGGAC	GCGCGTGGAA
	751	AAATTCGGCA	CCGGACACGG	CTGCGAGATT	GGTGAAGTCA	ATGCGACGG
	801	CTCGTTCGAC	GTAATTCGTT	ACGGCAAA	AGCCGGACAC	GTGCGATGGG
	851	ATTTCGATGG	CGGACACAA	CGCATGAAC	CGCTCGCGGT	CATCGCTGCC
40	901	GCACGCCATG	CCGGAGTCGA	TGTTCAAGC	GCTTGGAAG	CCTTGGGTGC
	951	GTTTAAAAAC	GTCAAACGCC	GCATGGAAT	CAAAGGCACG	GCAAAACGGCA
	1001	TCACCGTTTA	CGACGATTTC	GCCCAACAC	CGACGCCAT	CGAAACACAG
	1051	ATTCAAGGTT	TGCGCCACAG	TGTCGGGCGC	GCGCGCATCC	TGCGCGTCT
	1101	CGAGCGCGGT	TCCACACCA	TGAAACTCG	CACGATGAAG	TCCGCGCTGC
	1151	CCGCAAGCCT	CAAAGAACCC	GACCAAGTGT	TCTGCTACGC	CGCGCGCGCG
45	1201	GACTGGGACG	TGTCGGAAGC	CCTCGCGCGT	TGGGCTGCA	GCGTGCAGGT
	1251	CGGTAAGAT	TTTGATACCT	TGCTTGCCGA	AATTTGTGAA	AACGCCCGGA
	1301	CGCGCGACCA	TATTTTGGTG	ATGAGCAAC	GCGGTTTCGG	CGGAATACAC
	1351	ACCAAACTGC	TGACGCTTT	GAGATAG		

This corresponds to the amino acid sequence <SEQ ID 874; ORF132ng-1>:

55	1	MKHHIIIGIG	GTFMGGIAAI	AKEAGFKVSG	CDAKMYPPMS	TQLEALGIGV
	51	HEGFDAQAQLE	EFQADIYVIG	NVARRGMDVV	EAILNRGLPY	ISGFWLAENV
	101	VLHHHVLGV	AGTHGKTTTA	SMIAWVLEYA	GLAPGFLIGG	VFGKFRFRFP
	151	LPPTPRODEN	SKSFFVIEA	DEYDTAFDK	RSKFVHYRPR	TAVLNNLDF
	201	HADIADLGA	IOTOFHILVR	TVPSGLIVC	NGOQSLQDT	LDKGWTPVE
	251	KFGTGHGWOI	GEVNADGSFD	VLLDGKAGH	VANDMLMGHN	RNALAVLIA
60	301	ARHAGVDVQT	ACEALGAFKN	VKRMEIKGT	ANGITVYDF	AHHPTAIETT
	351	IQLRLQRVGG	ARILAVLEPR	SNMTKLGTMK	SALPASLKEA	DOVFCYAGGA
	401	DWDVAELAP	LGCLRVRGKD	FDTFVAEIVK	NARTGDHILV	MSNGFGGGIH
	451	TKLLDALR*				

ORF132ng-1 and ORF132-1 show 93.2% identity in 458 aa overlap:

```

5  orf132ng-1.pep MKHHIIGIGGTFMGGIAAIAKEAGFKVSGCDAKMYPMMSTQLEALGIGVHEGFDAQAQLE
    orf132-1      MKHHIIGIGGTFMGGIAAIAKEAGFEVSGCDAKMYPMMSTQLEALGIDVYEGFDDAQAQLD
10  orf132ng-1.pep EFQADIYVIGNVARRGMDVVEAILNRGLPYISGPQWLAEVNLHHHWWLVGAGTHGKTTTA
    orf132-1      EFQADYVIGNVARRGMDVVEAILNRGLPYISGPQWLENVNLHHHWWLVGAGTHGKTTTA
15  orf132ng-1.pep SMLAWVLEYAGLAPGFLIGGVPENFVSGARLPQTPRQDPNSKSPFFVIEADYDTAFFDK
    orf132-1      SMLAWVLEYAGLAPGFLIGGVPENFVSGARLPQTPRQDPNSQSPFFVIEADYDTAFFDK
20  orf132ng-1.pep RSKFVHYRPTAVLNNLEFDHADIFADLGAIQTFHHLVTVPSSEGLIVCNGQQQSLQDT
    orf132-1      RSKFVHYRPTAVLNNLEFDHADIFADLGAIQTFHHLVTVPSSEGLIVCNGQQQSLQDT
25  orf132ng-1.pep LDKGCWTPVEKFGTGHGWQIGEVNADGSDFDVLLDGKKAGHVAWDLMGHNRNMALAVIAA
    orf132-1      LDKGCWTPVEKFGTGHGWQAGEANADGSDFDVLLDGKTAGRVKWDLMGRHNRNMALAVIAA
30  orf132ng-1.pep ARHAGVDVQTACEALGAFKNVKKRMEIKGTANGITVYDDFAHHPHTAIETTTIQLRQRVGG
    orf132-1      ARHGVVDIQTACEALGAFKNVKKRMEIKGTANGITVYDDFAHHPHTAIETTTIQLRQRVGG
35  orf132ng-1.pep ARILAVLEPRSNMTKLGTMKSALPASLKEADQVFCYAGGADWDVAEALAPLGCRLRVGKD
    orf132-1      ARILAVLEPRSNMTKLGTMKSALPVSLKEADQVFCYAGGVWDVAEALAPLGCRLRVGKD
40  orf132ng-1.pep FDTFVAEIVKNARTGDHILVMSNGFGGIHTKLLDALRX
    orf132-1      FDTFVAEIVKNAEVDGHLVMSNGFGGIHTKLLDALRX

```

In addition, ORF132ng-1 is homologous to a hypothetical *E. coli* protein:

```

35  pir||856459 hypothetical protein o457 - Escherichia coli >gi|537075 (U14003)
    ORF_o457 [Escherichia coli] >gi|1790680 (AE000494) hypothetical 48.5 kD protein
    in fbp-pmba intergenic region [Escherichia coli] Length = 457
    Score = 474 bits (1207), Expect = e-133
    Identities = 249/439 (56%), Positives = 294/439 (66%), Gaps = 13/439 (2%)
40  Query: 22 KEAGFKVSGCDAKMYPMMSTQLEALGIGVHEGFDAQAQLEEFQADIYVIGNVARRGMDVVE 81
    Sbjct: 21 RQLGHEVTGSDANVYPMSTLLEKQGIETLQSYDASQLEP-QPLDVIGNAMTRGNPCVE 79
45  Query: 82 AILNRGLPYISGPQWLAEVNLHHHWWLVGAGTHGKTTTASMLAWVLEYAGLAPGFLIGGV 141
    A+L + +PY+SGPQWL + VL VWL VAGTHGKTTTA M +*LE G PG+*IGGV
    Sbjct: 80 AVLEKNIPIYMSGPQWLHDFVLRDRVWLVAVAGTHGKTTT*AGMATWILEQQGVKPGFVIGGV 139
50  Query: 142 PENFGVSARLPQTPRQDPNSKSPFFVIEADYDTAFFDKRSKFVHYRPTAVLNNLEFDH 201
    A+L + +PY+SGPQWL + VL VWL VAGTHGKTTTA M +*LE G PG+*IGGV
    Sbjct: 140 PGNFEVSAHL-----GESDFFVIEADYDCAFFDKRSKFVHYCPRTLILNNLEFDH 190
55  Query: 202 ADIFADLGAIQTFHHLVTVPSSEGLIVCNGQQQSLQDTLDKGCWTPVEKFGTGHGWQIG 261
    ADIF DL AIQ QFHHLVR VP +G I+ +L+ T+ GCW+ E G WQ
    Sbjct: 191 ADIFDLKAIQKQFHHLVTVPSVPGGRIIWPENDINIKQTMAGCWSQELVGEQGHWOAK 250
60  Query: 262 EVNADGS-FDWLLDGKKAGHVAWDLMGHNRNMALAVIAAARHAGVDVQTACEALGAFKN 320
    ++ D S ++VLLDG+K G V W L+G HN N L TAAARH GV A ALG+P N
    Sbjct: 251 KLTTDASEWEVLLDGEKGVGVKSLVGEHNHNLMAIAARHGVGAPADAANALGSFTN 310
65  Query: 321 VKRMEIKGTANGITVYDDFAHHPHTAIETTTIQLRQRVGG-ARILAVLEPRSNMTKLGTM 379
    +RR+R++G ANG+TVYDDFAHHPHTAI T+ LR +VGG ARI+AVLEPRSNMTK+G
    Sbjct: 311 KRRLELRGEANGVTVYDDFAHHPHTAILATLAALRGKVGCTARIIVLEPRSNMTMGIC 370
70  Query: 380 KSALPASLKEADQVF-CYAGGADWDVAEALAPLGCRLRVGKDFDTFVAEIVKNARTGDH 438
    K L SL AD+VF W VAE D DT +VK A+ GDHI
    Sbjct: 371 KDDLAPSLGRADEVFLQPAHPIWQVAEACVQPAHWSGDDVTLADMVVKTAQPGDHI 430
    Query: 439 LVMSNGFGGIHTKLLDAL 457

```

LVMSNGGFGGIH KLLD L
 Sbjct: 431 LVMSNGGFGGIHQKLLDGL 449

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

- 5 ORF132-1 (26.4kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 20A shows the results of affinity purification of the His-fusion protein, and Figure 20B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for FACS analysis (Figure 20C) and ELISA (positive result). These
 10 experiments confirm that ORF132 is a surface-exposed protein, and that it is a useful immunogen.

Example 103

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 875>

```

1  ..CCGGCGTATT ACGGCTCGGA TGACGAATTT AAGCGGCGAT TCGGAGAAAA
15  51  CTGCGCGACA TmCAGAAGAC ATTGCAACCG GAGCTGCGGG ATTTATGAA
101  CCGTATTGAA AAAATACGCG AAAAAGCGCG CAACACACCA TTCGGTCAGC
151  ATTAGTCGGG ACTTGGCGGA TATTTCATG CCCTTGGCGA GCTATTGCGG
201  CACACACCGT ATGCCCAACA TCCAGAAGAT GTATTTTTCC CAATTCGGCG
251  ACTCGCGCGT TCACACCGCC TTAAACACAG AGCGCGCAAA CACTTGGCAA
301  TTTGGCTTCi ATACCTATAA AAAAGGATTG TTAACACAG ATGATACATT
20  351  AGGATTAAAA CTGGTCGGCT ACGCGACGCG CATCGACAA TACATCCACA
401  ACGTTTACGG GAAATGGTGG GATTTGAACG GGTATTTCC GAGCTGGGTC
451  AGCAGCAGCG GCCTTGCTTA CACCATCCAA CATCGCATT TCAGAGACAA
501  AGTCGATCAA nnnnnnnnnn nnnnnnnnnn nnnnTACGAT TATGGCGGTT
551  TTTTACACAA CCTTCTTTAC GCCTATCAAA AAGCAGCGCA ACCGACCAAC
25  601  TTCAGCGATG CGAGCGAATC GCCCAACAA GCGTCCAAAG AAGACCAACT
651  CAACCAAGGT TATGGGTTGA GCAGGGTTTC CGCCTGCGG CGAGATTACG
701  GACGTTTGGG AGTGGGTACG CGCTGGTTGG GCAACAAATC GACTTTGGGG
751  GCGCGGATGC GCTATTTCGG CAGAGCATC CGCGGACGG CTGAAGAACG
30  801  CTATATCGAC GGCACCAACG GGGGAATAT CAGCAATTTT CGGCAACTGG
851  GCAGCGGTTT CATCAACACA ACCGAACCTC TTGCCCGCCA GCCTTTGATT
901  TTGATTTTAA ACGCGCGTTA CGAGCGGAG AAAAACCTTA TTTTCGCG
CGAGTCAAA ATCTGCTTCG ACAGCGCTTA TATCGATCGC CTCGATCGCG
101  GCATGATCG GGCACAGC GGTATTACCA CTGCTGCGA CCGCAAGAC
1051  AAGCACTTAC AGCTAACGTG TATGCTGAT AAAAGTTGTT GCaACGGCAA
35  1101  ATACCGCGCG CAAAGCAAAA GCGTATTGAC CAATTTTGA CGCGACGCGA
1151  CCTTTTgAT GAOGATGAGC TACAAGTTTT AA
```

This corresponds to the amino acid sequence <SEQ ID 876; ORF133>:

```

1  ..PGYYGSDDEF KRAFGENSPT KKKHCNRSCG IYEPVLKKYV KKRANNHVS
15  51  ISADFQDYFM PFASYSRTHR MPNIQEMYSF QIGDSGVHTA LKPERANTWQ
40  101  PGFXTYKRG LKQDDTLGLK LVGYRSRID YIHNVYKMW DLNGDIPSW
151  SSTGLAYTIQ HRXFXDKVHQ XXXXXXXXYD YGRFFTNLW AYQKSTOPTN
201  FSDASESPNN ASKEDQLKQG YGLSRVSALP RYGRLEVGT RNLGNKLTIG
251  GAMRYFGKSI RATAERYID GTNNGNTSNF RQLGKRSIQ TETLARPLI
30  301  XDFNAAYEPK KNLIFRAEIV NLFDRRTYID LDAGNDAAKE RYSSFPDPKD
45  351  KDXDVTICAD KTLNCKYGG TSKSVLTNFA RGRTEMTMS YKF*
```

Further work revealed the further partial DNA sequence <SEQ ID 877>:

```

1  GAGGCGCAGA TACAGTTTTT GGAAGATGTG CACGTCAAGG CGAAGCGCGT
15  51  ACCGAAAGAC AAAAAAGTGT TTACCGATGC GCGTCCGCTA TCGACCCGTC
101  AGGATATATT CAATCCAGCG GAAAACCTCG ACACACTCGT ACGCAGCATC
50  151  CCGGTCGGGT TTACACAGCA AGATAAAGC TCGGCGATG TGTCTTGA
201  TATTCGCGGC GACAGCGGTT TCGGCGGGGT CAATACGATG GTGACGCGCA
251  TCACGCGAGC CTTTATTTG ACCTTACGCG ATCGGCGCAG GCGAGCGCGT
```

301	TCATCTCAAT	TGGTGCATC	TGTGCACGC	AATTTTATTG	CCGAGCTGGA
351	TGTCTCAAA	GGCAGCTTCA	GGGCTCGCG	AGGCATCAAC	AGCCTTGCCG
401	GTTGGCGGAA	TCTGCGGAGT	TTAGCGCTGG	ATGACGTCGT	TCAGGCGCAAT
451	AATACCTACG	GCCTGCTGCT	AAAAGTCTG	ACCGGCACCA	ATTCAACCAA
501	AGGTAAATGG	ATGGCGGCGA	TAGGTGCGCG	CAAAATGGCTG	GAAGCGGAG
551	CATCTGTGCG	TGTGCTTTAC	GGGCACAGCA	GGCGCAAAAT	GGCGCAAAAT
601	TACCGCGTGG	GCGGCGGCGG	GCAGCAGCATC	GGAAATTTTG	GCGCGGAATA
651	TTTGGAAACG	CGCAAGCAGC	GATATTTTGT	ACAAGAGGTT	GCTTTGARAT
701	TCAAATTCGA	CAGCGGAAAA	TGGGAGCGGG	ATTTACAAGG	GCAACAGTGG
751	AAATACAGCG	CGTATAAAAA	TTACAACAAC	CAAGACTATC	AAATAATACAT
801	CGAAGAGCAT	GACAAAGCTT	GGCGGAAAAA	CCTg, CaCG	CAATACGACA
851	TTACCCCATC	CGATCCGCTC	AGCCTGAAGC	AGCAGTCGGC	AGGCATCTCG
901	TTTAAATTTG	AATACGACGG	CGTATTCAAT	AAATACACGG	CGCAATTTTG
951	CGATTTAAAC	ACCAAAATCG	GCAGCSCGAA	AATCATCAAC	CGCAATTTATC
1001	AGTTCAATTA	CGGTTTGTCT	TTGAACCGGT	ATACCAACCT	CAATCTGACC
1051	GCAGCTACA	ATTTCGGCAG	GCAGAAATAT	CGAAAGGGT	CGAATTTTAC
1101	AGCGTGGGG	CTTTTAAAGG	ATTTTGAAC	CTACACACAC	GGGAAATTC
1151	TCGAGCTCAA	CANACGCGCC	ACCTTCGCGC	TGCGCCGCGA	AACCGAGTTG
1201	CAAAACACTT	TGGGCTTCAA	TTATTTCCAC	AACAGATACG	GCAAAAACCG
1251	CTTTCTGTAA	GAATTTGGGG	TGTTTTTCGA	CGGTCCTGAT	CAGGACAACG
1301	GGCTTTATTCT	CTATTTGGGG	CGGTTTAAAG	GOGATTAAGG	GCTGCTGCC
1351	CAAAAATCAA	CCATTTGTCCA	ACCGGCGCGG	AGCCAAATAT	TCMAACGGTT
1401	CTACTTCGAT	GCGCGGCTCA	AAAAAGACAT	TTACGCGTTA	AACTACAGCA
1451	CCAATACCGT	CGGCTACCGT	TTGCGGCGCG	AAATATACGG	CTATTACCGG
1501	TCCGATGACG	AAATTTAAGCG	GGCATTCGGA	GAATACTCGC	CGCATACACA
1551	GAACATTGTC	AAACCGGAGCT	GCGGGATTTA	TGAACCCGTA	TTGAAAAAAT
1601	ACGGCAAAAA	GCGCGCCAAC	AACCAITCGG	TCAGCATTAG	TGCGGACTTC
1651	GGCGATTATT	TCATGCCGTT	CGCCAGCTAT	TGCGGCACAC	ACCGTATGCC
1701	CAACATCCAA	GAATGTATT	TTTCCAAAT	CGGCGACTCC	GGCGTTCACA
1751	CGCGCTTAAA	ACCAGAGCGC	GCAAAACACTT	GGCAATTTTG	CTTCAATACC
1801	TATAAAAAAG	GATTTGTTAAA	ACAAGATGAT	ACATTAGGAT	TAAACCTGGT
1851	CGGCTACCGT	AGCCGATCGC	ACAACATCAAT	CCACACAGTT	TACCGGAAAT
1901	GGTGGGATT	GACCGCGGAT	ATTCCGAGCT	GGGCTACAGC	CAACCGGCTT
1951	GGCTACACCA	TCCAAACATG	CAATTTCAAA	GACAAAGTGC	CAACAAACGG
2001	TTTTGAGTTG	GAGCTGAAT	AGATATATGG	GGGTTTTTTC	ACCAACCTTT
2051	CTTACGCCCTA	TCAAAAAGC	AGCACAACGA	CCAACCTTAC	CGATGCGAGC
2101	GAATCGCCCA	ACAAATGGCT	CAAAGAAGAC	CAACTCAAC	AAGGTTATGG
2151	GTTGAGCAGG	GTTTCGCGCC	TGCGCGGAGA	TTACGCGAGT	TTGGAAGTCG
2201	GATCGCGCTG	GTTGGCGCAAC	AACTGACTTT	TGGGCGGCGC	GATGCGCTAT
2251	TTCCGCAAGA	GCATCGCGCG	GACGGCTGAA	GAACGCTATA	TCGACGGCAC
2301	CAACGGGGGA	AATACAGCAG	ATTTCGCGCA	ACTGGGCAAG	CGTTCCATCA
2351	AACAAACCGA	AACCTCTGCC	CGCCAGCTTT	TGATTTTTGA	TTTTTACGCC
2401	GCTTACGAGC	CGAAGAAAAA	CCTTATTTTC	CGCGCGGAGG	TCAAAAATCT
2451	GTTTCGACAG	CGTTATATCG	ATCGCTCGA	TGCGGGCAAT	GATGCGGCAA
2501	CGCAGCGTTA	TTACAGCTCG	TTGACACCGA	AAGACAGGA	CGAAGACGTA
2551	ACGTGTAATG	CGTATAAAC	GTTGTGCAAC	GGCAAAATCG	GCGGCACAG
2601	CAAAAGCGTA	TGACCAATT	TTGACGCGG	AGCACCTTT	TTGATGACGA
2651	TGAGCTACAA	GTTTTAA			

This corresponds to the amino acid sequence <SEQ ID 878; ORF133-1>:

50	1	EAQIQWLEDV	HVKAKRVPKD	KKVFTDARAV	STRQDIFKSS	ENLNDIVRSI
	51	PGAFITQDQS	SGIVSLINRG	DSGFGRWNHM	VDGTQTFYFS	TSTDAGRAGG
	101	SSQFGASVDS	NFIAGLDVVK	GSFSSAGIN	SLAGSNILRT	LGVDVVQGN
	151	NTYGLLLKLG	TGNTSTKGNA	MAIAGRKWL	ESGASVGVLV	GHRSRSVAQN
	201	YRVGGGQOHI	GNFGEAYLER	RKQRYFVQBG	ALKFNSDSGK	WERDLRQQW
55	251	KYKPYKNYNN	QELQKYEIEH	DKSWRENLPK	QYDITPIDPS	LKQQSAGNLS
	301	FKLEYDGVFN	KYTAQFRDLN	TKIGSRKILN	RNYQFNYGLS	LNPTYNLNL
	351	AAVNSGRQKY	PKGSKFTGWL	LKDFETYNN	AKILDLNNTA	TRFLPRETEL
	401	QTLTGPNYFH	NEYGKNRFPE	ELGLFFDGD	QDNGLYSYLG	RFKGDGKLLP
60	451	QKSTIVQFAG	SQYFNTFYFD	AALKKDIYRL	NYSTNTVGYR	FGGEYTGYYG
	501	SDDEFKRAFQ	ENSPTYKKHC	NRSCGIYEPV	LKKYQKKRAN	NHSVTSIADF
	551	GDYFMFPAFS	SRTHRPMNIQ	EMFYSQIGDS	GVHTALKPER	ANTWQPGENT
	601	YKKGLLQKDD	TGLGLKLVGYR	SLRINYLNH	YKGNWDLNGD	IPSWVSTSLG
	651	AYTTQHRNRF	DKVHKHGFEL	ELNIDYGRFF	TNLISYAYQKS	TOPTNFSAS
	701	ESPNNAKSED	QLKQGYGLSR	VSALEPRDYG	LEVGTBRLGN	KILTLGAMRY
65	751	FGKSIIRATAE	ERYIDGNTGG	NTSNFRQLGK	RSIQKQETLA	RQPLIFDEYA
	801	AYEPKKNLIF	RAEVKNLFDR	RYIDPLDAGN	DAATQRYYSF	FDPKDKDEDEV
	851	TCNADETLGN	GKYGGTSKSV	LTMFARGRTF	LMTMSYK*	

Computer analysis of this amino acid sequence gave the following results:

Homology with with the probable TonB-dependent receptor HI121 of *H.influenzae* (accession number U32801)

ORF133 and HI121 show 57% aa identity in 363aa overlap:

	Orf133:	31	ITYEPLVKLKQKRRANNHSVISAADFVYPMFPASYSRTHRMPNIQEMVSQLIGDSGVHTA	90
5			I EP+L K K G K A N H S +A+D+Y D Y M P F +Y S R T H R M P N I Q E M +F S Q + + G V +T A	
	Hi121:	563	INEPILHKSGHKKAFNHSATLSAELSDYFMPFFYSRTHRMPNIQEMFFSQVSNAGVNTA	622
	Orf133:	91	LKPERANTWQFGFXTYKKGLLQDDTLGLKLVGYSRIDNYHNVGVKQWINDGIPSWV	150
			LKPE+++TQ G F TYK L Q D D L G +K L V G Y S I N Y I H N V G V W W +P+W	
10	Hi121:	623	LKPEQSDTYQLGFNTYKKGLETQDDVLGVKLGVYSFIKNIYHNVGVVW---RDGMETWA	680
	Orf133:	151	SSGLAYTIQHRKFXDKDXXXXXXXXYDGRFFNTLSYAYQKSTQPTNFDASESPNN	210
			S G Y T I + + V Y D G R F N N +S Y A Y Q + + Q P T N +D A S P N N	
	Hi121:	681	ESNGFKYTIHAHQNYKPIPVKKSGVLEINYDMGRFFANVSYAYQRTNQPTNYADASPRPNN	740
15	Orf133:	211	ASKEDLKQCYGLSRVSLPRDYCRLEVGCTRWLGNKLTIGGMARYFGKSRIRATAEERYD	270
			AS+E D L K Q C Y G L S R V S L P R D Y C R L E V G C T R W L G N K L T I G G M A R Y +G K S R A T E E Y I +	
	Hi121:	741	ASQEDILKQCYGLSRVSMPLPKDYGRLELGTWFRDQKLTILGLAARYYGKSKRRATIEEYIN	800
	Orf133:	271	TGNGGNTSNFRQLGKRSIKOTETLARQLIXDFNAAVEPKKNLIFRAEKNVLFDRRYIDF	330
20			G + + R + + +K +T E + + Q E + I + +Y E P K +L I +A E V +N L +H R +D P	
	Hi121:	801	GSR-FKNTLRLRENYAVAKTDEDKKQFIILDLHVYFEPKIDLIKAEVQNLDKRYVDP	859
	Orf133:	331	LDAGNDAKERYSSDFPKDKDDTQV---CNADKLTLCNGYCGTSSKSLVTLNFARGRTFLMYS	390
			L D A G N D A A +R Y Y S S + + C D C D C G G + K +V L N F A R G R T + + + + +	
25	Hi121:	860	LDAGNDAASQRYSSYL-----NNSIECAQDSSAC---GGSKDVTLYNFAARGRTYILSIN	910
	Orf133:	391	YKF 393	
			YKF	
	Hi121:	911	YKF 913	

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF133 shows 90.8% identity over a 392aa overlap with an ORF (ORF133a) from strain A of *N.*

meningitidis:

[illegible]

10

15

20

25

30

35

40

45

50

55

60

65

1	A	AAGACAAA	AAGCTGTTTAC	CGATCGCGGT	CGCGTATGCG	CCGCTACGGA
2	T	TATTTCAAA	TACNCANGAA	ACCTCGACGA	CATCTGATGCG	ACNATCCCGC
101	G	TGGCGTTTAC	ACANCAANAT	AAAGCTCGGG	CGTTTGTGTC	TTTGAATATC
11	C	CGNGCGGCG	CGGGGTTCGG	CGGGTCGATC	ACNATGGTNG	AGCGCATCAC
201	N	NCANACATCT	TTATCGCACT	TACCGGATGC	CGGACGGCGA	CGGGGTTTAT
251	C	CTCAATTCCG	TGCATCTGTC	GACAGCATCT	TTGTCGCGGG	ACTCGATGTAT
301	G	CTCAAAGGCA	CGTTCAGCGC	CTGCGGAGCG	ATCAACACGCG	TTCGCGGTTG
351	G	CGCGATCTCG	CGGACTTTAN	CTGTGGATGA	TCGTGTTTCC	TGCATANTTA
401	C	NTACGCGCCT	CGGTCTAAAA	GGTTCGACCG	CGACCAATTC	ACGCAAAAGT
451	A	AATGCGATGG	CGGCGATGAG	TGCGCGACAA	TGGCTGGAAA	TCGGCGATCC
501	T	TGTGGGTGTC	CTTTACGGCG	ACAGCAGCGG	CAGCGTGGCG	CGAAATATCC
551	C	CGTGCGGGCG	CGGTGGCGGAG	CAGCATCGAA	TTTGGGCGCG	GGAATATCTG
601	G	CAAGCGCGCA	ACGCRACGAT	TTTGGAGCAA	GAGGGGCGTG	TGAATCGAAT
651	C	CGGCTGCGCT	CGGCGGCGCG	CGGAGGCGCG	CGAAGGCGCG	CGGCGGCGCG
701	C	CGACTGGTGA	TCAAATAATC	GATCGCCCCC	AGAAATCTCA	AAAATACATC
751	G	GAAGTCATG	ATAAAGACCG	CGCGGAARAC	CTGGCGCGCG	ATAACGACAT
801	C	CACCCCCATC	CGTCGCTCCA	CGTATGAAGN	CGATCGCGCA	CGCAATCCGT
851	T	TAAATTTGGA	TACAGCGCGC	GTATTCATCA	ATAACACGCG	CGAATTTCCG
901	G	GAATTAACAA	CCAAATACGG	CAGCGCGAAA	ATCATCACCC	CGAATTAATC
951	A	ATTCATATAC	GTTTTTGCTT	TGACACCGTA	TACCAACCTC	ATATCTGACG
1001	G	CGCATACAAA	TTCGGGCGAG	CGAARAATTC	GAGAGGGTGC	GAGTTTTCAC
1051	G	GGCTGGGGCG	TTTNNRAAG	TTTGAAGCTT	TACACACACG	CAAAAATCTG
1101	C	CGCTCATCNC	ACCACTCCCA	CTTTCGACCG	CGCCGGTAGC	ACCGAGTTGC
1151	T	AAACCACTTT	GGGCTTCAAT	TATTTCCACA	ACGAAATACG	CAAAAACCGG
1201	T	TTTCTGAAG	AAATGGGGCT	GTITTTTCAG	CTGCGGGTAT	ANGACAACCG
1251	G	CGTTTATTC	TATTTGGGCG	GGTTTAAAGG	CGATTAAGGG	CTGTCTGCCC
1301	A	AAAATATCAC	CTTTGTCCAA	CGGCGCGCGA	GCCAATATTT	CACAGCTGTC
1351	T	TACTTGATG	CGGCGCTCAA	AAAGAATGAT	TACCGCTTGA	ACTTACAGAC
1401	C	CAATACGCT	CGGTACGGTT	TGCGCGGCGA	ATATATGCGG	ATTATNCGCT
1451	A	CGGATGACGA	ATTTAAGCGC	CGATTCGAGT	AAATCTCGCC	GCATCTACAG
1501	T	GAATTTGCA	CGGCGGCGCG	CGGACGATAT	ACACATTCAT	CGGCGGCGCG
1551	C	CGGATTTGCA	CGCGACACAA	CGGACGATAT	CGGCGACGAT	CGGAGCTTAT
1601	N	CGGATTTGAT	CAATCGCTTC	CGGACGATTT	CGGCGACACA	CGGTTACACG
1651	A	CAATTTCAAG	CAATGTATTT	TTCCCAATAT	CGGCGATCCG	CGGCTCGCCG
1701	G	CGCTTTAAAA	CCAGGCGCGG	CAACAATCTG	CGAATTTGGC	TTAAATCACT
1751	A	ATAAARAAG	ATTGTTAAGG	CAGATGATTA	TATTAGGATAT	CAAAATGGTC
1801	G	GGCTACGCGA	CGCGCATGCA	NCTACATCTC	CACACAGTTT	ACCGGGAATG
1851	T	GGGGATTG	AACGGGAATA	TTCCGAGTCT	CGGACGAGCG	ACCGGGCTTG
1901	T	TTGACCATCT	CCRAACACCG	AAATTTCAAG	ACAAAGTGCA	CAACAACGCT
1951	T	TTTAGTTTGC	ACGCTGAATTA	GAAATTTCAG	CGTTTTTTCG	CAACACTTTC
2001	T	TTAGCGCTAT	CAAAAAGAACA	CGACACCGAC	CAAACTTCAGC	GGTGCGAGCG
2051	A	AATCGGCCAA	CATCGCTGTC	RAAGAAGACC	AACCTCAACCA	AGATTTATGG
2101	T	TGCGGAGGG	TTTTCGCGCT	CGCGCGAGAT	TACGAGGCTT	TGGAGCTCGG
2151	T	TAGCGCGTTC	TTGGGCAACA	AACCTGATAT	GGGCGCGCGG	TCGCGCTAT
2201	T	TCGGCAGGAG	CATTCGCGCG	ACGGCTGAAG	ACGCTATATC	CGAGNCACCC
2251	A	AATGGGAGNN	NTACACGCAA	TTTCGGGCAA	CTGGGCGAAG	TCTTCATCAN
2301	C	ACRAACGCAA	ACCTTTGCGC	CGACGGCTCT	GATTTTGTAT	TNTNACGCT
2351	T	CTTAGCGAGT	GAGAAAANAAT	CTATATTTTC	GGCGCGAGAT	CAAAAATCTG
2401	T	CTTAGCGAGT	CGGCGGCGCG	TTCGCGAGAT	CGGCGCGGAT	CGGCGGCGCG
2451	G	GGAGTATGCA	CGCACTTGCT	CGGCGGCGCG	CGACAGACAG	GAGAACTGAA
2501	G	TGCTTATGTA	GTATAACACG	TTTATGACAG	GCRAATACGG	CGGCAACGCA
2551	A	AAAGCGTAT	TGACCAATTT	TGCACGCGCA	CGCACTTTTT	TGATACGATAT
2601	G	GAGCTACAG	TTTTTAA			

This encodes a protein having (partial) amino acid sequence <SEQ ID 880>:

```

      1 KDKKVFTDAR AVSTRQDIFK SXENLDNIVR XIPGAFTXQK KSSGXVSLNI
      51 RXDSGFGFRVN TMDVGTXTFT YSTSTDAGRA GGSQFGASV DSNFXAGLDV
      101 VKGSFSGSAG INSLAGSANL RTLKVDVVQ GNXTYGLLLK GLTGTNSTKG
      151 NAMAAGIARK WLESGASGVG LYGHSSRSVA QNYRVGGGGQ HIGNFGAEYL
      201 ERRKQRYFEQ EGGLKFNSSN GKWERDFQKS YWTKWYQKY DAPQELQKYI
      251 EGHDKSWREN LAFQYDITPI DPSSLLKQSA GNLFKLEYDG VNKYTAQFR
      301 DANKTKGRK TIIRNYPQNY GLSNPNTNL NLTAAVNSGR QYKPKGSKT
      351 GWGLXKDFET YNNAKILDLL NTSTFRLPRE TELQTTIGFN YEHNEYGNR
      401 FPEELGLFDP GPDXDNGLYS YLGRFKGDKG LLPQKSTIVQ PAGSQYNTF
      451 YFDAALKKDI YRLNYSNTV GYRFGCYTG YXXSDDFKR AFGENSPYTX
      501 KHCNQSQGIY EPVLKKYQKK RANNHVSVIS ADFGDTFMPF ASYSRTHRMP
      551 NIQEMYFSQI GDSGVHTALK PERANTWQFG FNTYKGLLK QDDILGLKLV
      601 GYRSRIDXYI HNVYKWWDL NGNIPSWVSS TGLAYTIQHR NFKDKVHKHG
      651 FELELNIDYX RFFTNLSYAY QKSTOPTNFS DASEPNNAS KEDOLKQYV
      701 LSRVVSALPRD YGRLEVGTW LGNKLTLGGA MRYFGKSIRA TAEERYIDXT
      751 NCXXTSNFRQ LGRKSIXQTE TLAQPLIFD XYAAYPEKKX LIFRAEVKNL
      801 FDRRYIDPLD AGNDAATQRY YSDFDPKDKD EEVTCNDNDT LCNGKYGCTS
      851 KSVLTNFARG XTFLITMSYK F*
  
```

20 ORF133a and ORF133-1 show 94.3% identity in 871 aa overlap:

```

      10      20      30      40
or133a.pep      KDKKVFTDARAVSTRQDIFKSXENLDNIVRXIPGAFTXQKKS
25 or133-1      EAQIQVLEDVHVHAKARVPEKDKVFTDARAVSTRQDIFKSENLDNIVRSTPGAFTQQQKS
      10      20      30      40      50      60
or133a.pep      SGXVSLNIRKXDSGFRVNTMTVDGITXTFYSTSTDAGRAGGSQFGASVDSNFXAGLDVVK
30 or133-1      SGIVSLNIRGDSGFRVNTMTVDGITQTYSTSTDAGRAGGSQFGASVDSNFIAGLDVVK
      70      80      90      100      110      120
or133a.pep      GSFSGSAGINSLAGSANLTLXVDVVQGNXTYGLLLKGLTGNSTKGNAMAAGARKWL
35 or133-1      GSFSGSAGINSLAGSANLTLGVDDVVQGNNTYGLLLKGLTGNSTKGNAMAAGARKWL
      130      140      150      160      170      180
or133a.pep      ESGASVGVLYGHSRVSVAQNYRVGGGGQHIGNFGAEYLERRKQRYFEQEGGLKFNSSGK
40 or133-1      ESGASVGVLYGHSRVSVAQNYRVGGGGQHIGNFGAEYLERRKQRYFVQEGALKFNSSGK
      190      200      210      220      230      240
or133a.pep      WERDFQKSYWKTKWYQKYDAPQELQKYIEGHDKSWRENLAPOYDITPIDPSSLLKQCSAGN
45 or133-1      WERDLQRCQWKYPYKNYNN-QELQKYIEEHDKSWRENLXPOYDITPIDPSSLLKQCSAGN
      250      260      270      280      290
or133a.pep      LFKLEYDGVFNKYTAQFRLNTKIGSRKIINRNYQFNHGLSLNPTTNLNLTAAYNSGRQK
50 or133-1      LFKLEYDGVFNKYTAQFRLNTKIGSRKIINRNYQFNHGLSLNPTTNLNLTAAYNSGRQK
      300      310      320      330      340      350
or133a.pep      YPKGSKFTGWGLXKDFETYNNAKILDNLNTSTFRLPRETELQTTIGFNHYEHNEYGNRFP
55 or133-1      YPKGSKFTGWGLXKDFETYNNAKILDNLNTATFRLPRETELQTTIGFNHYEHNEYGNRFP
      360      370      380      390      400      410
or133a.pep      EEGLFFDFDPDXDNGLYSYLGRFKGDKGLLPQKSTIVQPAGSQYNTFYFDAALKKDIYR
60 or133-1      EEGLFFDFDPQDNGLYSYLGRFKGDKGLLPQKSTIVQPAGSQYNTFYFDAALKKDIYR
      420      430      440      450      460      470
  
```

		470	480	490	500	510	520
	orf133a.pep	LNVSTNTVGYRFGGXYTGYXSDDEFKRAFGENSPYTKHCNCSGIEPVLKKYGKKRA					
5	orf133-1	LNVSTNTVGYRFGGXYTGYXSDDEFKRAFGENSPYTKHCNCSGIEPVLKKYGKKRA	480	490	500	510	520
		530	540	550	560	570	580
	orf133a.pep	NNHSVSISADFGDYFMPFASYSRTHRMPIQEMYSQIGDSGVHTALKPERANTWQFGFN					
10	orf133-1	NNHSVSISADFGDYFMPFASYSRTHRMPIQEMYSQIGDSGVHTALKPERANTWQFGFN	540	550	560	570	580
		590	600	610	620	630	640
	orf133a.pep	TYKGLLKQDDTLGLKLVGYRSRIDNYIHNVYQKWDLNGNIPSWVSTGLAYTIQHNR					
15	orf133-1	TYKGLLKQDDTLGLKLVGYRSRIDNYIHNVYQKWDLNGNIPSWVSTGLAYTIQHNR	600	610	620	630	640
		650	660	670	680	690	700
	orf133a.pep	KDKVKHKGFELELNVDYGRFFTNLSYAYQKSTOPTNFSASESPNNAKEDGLKQGYGLS					
20	orf133-1	KDKVKHKGFELELNVDYGRFFTNLSYAYQKSTOPTNFSASESPNNAKEDGLKQGYGLS	660	670	680	690	700
		710	720	730	740	750	760
	orf133a.pep	RVSALPRDYGRLEVGTWLGKNTLGGAMRYFGKSI RATAEERYIDXTNGXSTNPFQLG					
25	orf133-1	RVSALPRDYGRLEVGTWLGKNTLGGAMRYFGKSI RATAEERYIDXTNGXSTNPFQLG	720	730	740	750	760
		770	780	790	800	810	820
	orf133a.pep	KRSIKQTETLARQPLIFDXYAAYEPKKNLIFRAEVKNLFDRRYIDPLDAGNDAATQRYYS					
30	orf133-1	KRSIKQTETLARQPLIFDXYAAYEPKKNLIFRAEVKNLFDRRYIDPLDAGNDAATQRYYS	780	790	800	810	820
		830	840	850	860	870	
	orf133a.pep	SFDPKDKDEEVTGNNDNTLNGKYGGTSGSVLTNFARGXTFLMTSYKFX					
35	orf133-1	SFDPKDKDEEVTGNNDNTLNGKYGGTSGSVLTNFARGXTFLMTSYKFX	840	850	860	870	880

Homology with a predicted ORF from *N.gonorrhoeae*

ORF133 shows 92.3% identity over 392 aa overlap with a predicted ORF (ORF133ng) from *N.*

45 *gonorrhoeae*:

	orf133.pep	PGYYSDDDEFKRAFGENSPYTKHCNCSGIE	31
	orf133ng	FYFDAALKKDIYRLNYSNAINRYRFGGXYTGYGSENEFKRAFGENSPAYKEHCDPSGGL	560
50	orf133.pep	YEVLKKYGGKRRNNHSVSISADFGDYFMPFASYSRTHRMPIQEMYSQIGDSGVHTAL	91
	orf133ng	YEVLKKYGGKRRNNHSVSISADFGDYFMPFAGYSRTHRMPIQEMYSQIGDSGVHTAL	620
55	orf133.pep	KPERANTWQFGFYTKYKGLLKQDDTLGLKLVGYRSRIDNYIHNVYQKWDLNGNIPSWVS	151
	orf133ng	KPERANTWQFGFNFTYKGLLKQDDTLGLKLVGYRSRIDNYIHNVYQKWDLNGNIPSWVG	680
	orf133.pep	STGLAYTIQHRRFXDKVHQXXXXXXXXXYDGRFFTNLSYAYQKSTOPTNFSASESPNNA	211
60	orf133ng	STGLAYTIQHRRFKDKVHKHKGFELELNVDYGRFFTNLSYAYQKSTOPTNFSASESPNNA	740
	orf133.pep	SKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKNTLGGAMRYFGKSI RATAEERYIDG	271
	orf133ng	SKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKNTLGGAMRYFGKSI RATAEERYIDG	800
65	orf133.pep	TNGGNTSNFRLGKRSIKQTETLARQPLIXDFNAAAYEPKKNLIFRAEVKNLFDRRYIDPL	331
	orf133ng	TNGGNTSNVRQLGKRSIKQTETLARQPLIFDXYAAYEPKKNLIFRAEVKNLFDRRYIDPL	860

```

orf133.pep      DAGNDAAAXERYSSFDPKDKDXDVTCNADKTLGNGRYGGTSSKSVLTNFARGRTFLMTMSY 391
|||||
orf133ng        DAGNDAAATQRYSSFDPKDKDEDVTCNADKTLGNGRYGGTSSKSVLTNFARGRTFLMTMSY 920
|||||
orf133.pep      KF 393
||
orf133ng        KF 922

```

The complete length ORF133ng nucleotide sequence <SEQ ID 881> is predicted to encode a protein having amino acid sequence <SEQ ID 882>:

	1	MRSSRF.KPI	CFYLMGMVLY	HHSYAREDRG	AGSEAOIQVL	EDVHVKAKRV
	51	PKDKKVFTDA	RQSVLSTQDF	KSGENLADN	IRSPGAFTQQ	DKSSGIVRSTL
	101	IRGDSGFGRV	NPMWDGIGTQ	FYSTSTDADR	AGGSSQFGAS	VDSMFIAGLD
	151	VKGSFGSGSA	CNLSGASGN	LRTLGSTDVN	QDNNYTGILL	KGLTGTNSTK
15	201	GNAMAAIGAR	KWELSGASGV	VLYGHSRRGV	QDNTYVGGGG	QHIGNFGEFY
	251	LERRQQYFV	QEGGLKENAG	SKGWERDLQR	QWIKTWKWKY	YEDPQBLQYK
	301	IEEHDKSWRE	NLAOPYDITP	IPDSGLKQGL	AGNLNLNLEY	GVFNKYTAQF
	351	RDLNTRIGRS	KIINRYQFN	YGLSLNPYN	LNLTAAYNSG	RQKYPKGAKF
	401	TWGGLKDFE	TYNNAKILDL	NNTATFLRP	ETDQTLTLG	NYFHNSYGNK
20	451	RFPEELGLFF	DGPDQDNGLY	SYLGRKEGKL	KLFLQKSTIV	QPHAGSQFNT
	501	FYDAALAKD	IYRLNYSYNA	INYRFGEYET	GYGSGENEFF	RAFGENSPAY
	551	KEHCDPSGL	YEPVLKIKKY	KRANHNSVYI	SADPGDYFMP	FAGYSRTHRM
	601	PNIQMYRSDQ	IGDSGVHTAL	KPERANRQY	GFNTYKKGLL	KQDIDLGLKL
	651	VGYSRSDJNY	INYSYGRWMD	LANGDIFSWG	STGLATLRH	RNFKQDHHK
25	701	GFLLEYDYD	FTNLNLYA	YOKSTOPTNF	YKSTOPTNF	SKEDQLDGL
	751	GLSKVSALPR	DYGRLEKSTR	WLGNKLTLG	AMRYFGKSR	ATAEREYIDG
	801	TNGMSTNSR	QNLQKLVGRT	ETLARQPLIF	DIFYAEPFK	NILFRAEVKN
	851	LFDRRYIDPL	DAGNDATAOR	YYSFDFPKDK	DEDTVCNADK	TLCNGKYGGT
	901	SKSVI.TNFR	GRFTLMZMSY	KRT		

30 A variant was also identified, being encoded by the gonococcal DNA sequence <SEQ ID 883>:

	1	ATGAGATCTT	CTTCGCGTT	GAAAGCGATT	TGTTTTCATC	TTATGGGGTG
	51	TATGCTATAT	CATTACAGCT	ATGCGCAAGA	TGCGAGGACG	GCGGCGACGC
	101	AGGCGCAGAT	ACAGGTTTGG	GAAAGATTGC	ACGTCAAGCG	GAAAGCGCGTA
35	151	CCGAAAGACA	AAAAAGTGT	TACCGCTATG	CGTGCGCGTA	GCACCGCGTA
	201	gATGCTGTT	AAATCGGGCG	AAAAACCTCG	CAACATCGTA	CGCAGCATCT
	251	CGGGTGGCTT	TACACAGCAA	GATAAAGAG	GGCGGATTGT	GTCTTTGAAT
	301	ATTCGCGGGC	ACGCGGGTT	CGGGCGGGCT	AATACGATGG	TGGACGGCAT
	351	CACGACAGCC	TTTATTTGCA	CTTCTACGGA	TGCGGCGAGG	GACGCGGGTT
40	401	CATCTCAATT	CGGTCGATCT	GTGCGACAGA	ATTTTATTGC	GCGACTGGAT
	451	GTGCTCAAG	GCAGCTTCAG	CGGCTGGCGA	GGCATCAACA	GCCTTGGCGG
	501	TTGGCGGAAT	CTCGGAACTT	TAGGCGTGGA	TGACGTCGTT	CAGGGCAATA
	551	ATAACCTACG	CTCGTCTGTA	AAAGGCTGTA	CGCGACACAA	TTCAACACAA
	601	GTTATTCGGA	TGGCGGCGAT	AGGTGCGCGC	AAATGGCTGT	AAAGCGGAGC
45	651	GTCTGCTGGT	GTGCTTCACG	GCACGATCG	GCGCGCGTGT	GCGCGCGTGT
	701	ACCGGTCGCG	CGGCGCGGCG	CACCATCGCG	GAATTTTGGT	TGAAGAATAT
	751	TCGGAAACGC	GCAAAACAGA	ATATTTTGTA	CAAGAGGGTG	GTTTGAATTT
	801	CANTCGCGCG	ACGGGAAAT	GGGAACGGGA	TTTGAACGG	CAATACGGG
	851	AAACAAGATG	GTATAAAAA	TACGAAGACC	CCCAAGAACT	GCAAAATATC
50	901	ATCGAAGAGC	ATGATAAAG	CTCGGCGGAA	AACCTGGCGC	CGCAATA CGA
	951	CATCACCCCC	ATCGATCCGT	CGGCGCTGAA	GCAGCGTGGC	GCGAGCAATC
	1001	TGTTTAAATT	GGAATTACGAC	GGGCTATACA	ATAAATAAC	GGCGCAATTT
	1051	CGCGATTAA	ACACAGCAAT	CGGACGCGCG	AAAAATCATCA	ACCGCAATTA
	1101	TCGAATCAAT	TACGGTTTGT	CTTACGAC	TATACCAAC	CTCAATCTGA
55	1151	CGCGAGCCTA	CAATTGGGCG	AGGCAAGAA	ATTCGAAAGG	GGCGAAGTTT
	1201	ACAGGCTGGG	GGCTTTTAAA	AGATTTTAA	ACCTTACAACA	ACGCGAAAT
	1251	CTCGGACCTC	AAACAACACG	CACACTTTCG	GCTCGCCGCG	GAAACCGGAT
	1301	TGCAACAACC	TTTGGGCTCT	AATATTTCG	ACCAAGATTA	CGGCAAAAC
	1351	CGCTTCTCTG	AGGAATTGGG	GCGTGTTCCT	GACGGCTCTG	ATTCAGGACAA
	1401	CGGCGCTTAT	TCCTATTTTG	GGGCGGTTAA	GGGCGATAAA	GGGCTGTTCG
60	1451	CTCAAAATTC	AACCATGTCT	CACACGCGCG	GCAGCGCAATA	TTTCAACGCG
	1501	TTCTACTCTG	ATTCGCTGCG	CAAAAAGAC	TAAATACGCT	TAAATACGCT
	1551	CGCGTCGCA	ATCAACATCT	GTTCGCGGCT	GGAATATACG	GCGTATATCG
1601	1601	GCTCGGAAA	GAAATTTAAG	CGGGCTATCT	GGAAGAACTC	CGGCGCATCA
	1651	AAGGCAACT	GCGACCGGAG	CTCGGGCTCT	TATGACCGG	TATGACATAA
65	1701	ATACGCGCAA	AAGCGCGCCA	ACAACCAATC	GGTGCAGATT	AGTGGCGGAT
	1751	TGCGCGATTA	TTTCACTGCG	TGCTCGGCGT	ATTCGCGCAC	ACACCGTATG

1801	CCCAACATCC	AAGAAATGTA	TTTTTCCCAA	ATCGGCGACT	CCGGCGTTCA
1851	CACCGCCTTA	AAACCAAGAG	GCGCAACAC	TTGGCAATTT	GGCTTCAATA
1901	CCTATAAAAA	AGGATTGTTA	AAACAAGATG	ATATATTAGG	ATTGAAACTG
1951	GTGGCTACAC	CGAGCGCAT	TGACAACTAC	ATCCACAACG	TTTACGGGAA
2001	ATGGTGGGAT	TTGACCGGG	ATATTCCGAG	CTGGGTCGGC	AGCACC GGCG
2051	TTGCCTACAC	CATCCGACAC	CGCAATTCCA	AAGACAAAGT	GCACAAACAC
2101	GGTTTGAGC	TGGAGCTGAA	TACGATTAT	GGCGTTTTT	TCACCAACCT
2151	TTCTTACGCG	TATCAAAAAA	GCACGCAACC	GACCAATTTT	AGCGATGCGA
2201	GCGATCGGCC	CACACATCGC	tcacaaGAAG	ACCAACTCAA	ACAAGGTTAT
2251	GGGCTGAGCA	GGGTTTCCGC	CCTGCCGCGA	GATTACGGAC	GTTTGAAGT
2301	CGGTACGCGC	TGGTTGGGCA	ACAACTGAC	TTTGGCGGGC	CGGatgcGCT
2351	ATTTCCGCAA	GAGCATCCGC	GCGACGGCTG	AAGAAGCCTA	TATCGACGGC
2401	ACCAACGGGG	GAAATACCCG	CAATGTCGGG	CAACTGGGCA	AGCGTTCCAT
2451	CAACACAACC	GAAACCCCTG	CCCGACAGCC	TTGATTTTTT	GATTTTACG
2501	CCGCTTACG	GCGAGAGAAA	AACTTTATT	TCGCGCGGCA	AGTCAAAAC
2551	CTGTCCGACA	GCGCTTATTT	CGATCCGCTC	GATCGCGGCA	ATGATCGGCG
2601	AACGACGCGT	TATTACAGCT	CGTTCGACCC	GAAGACAAG	GACGAAGACG
2651	TACGCTGTAA	TGCTGATAAA	ACGTTGTGCA	ACGCGCAATA	CGCGCGCACA
2701	AGCAAAAGCG	TATTGACCAA	TTTCGCACGC	GGAAGCACT	TCTTGATGAC
2751	GATGAGCTAC	AAGTTTTAA			

This corresponds to the amino acid sequence <SEQ ID 884; ORF133ng-1>:

1	MRSSFRLKPI	CFYLMGVMLY	HHSYAEADGR	AGSEAQIQVL	EDVHVAKKRV
51	PKDKVFTDA	RAVSTRQDVF	KSENLENLIV	RSIPGFTQQ	DKSSGIVSLN
101	IRGDSGFRV	NTMVDGTTQT	FYSTSTDAGR	AGGSSQFGAS	VDSNFIAGLD
151	VVKGSFSGSA	GINSLAGSAN	LRTLGVDDVV	QGNNTYGLLL	KGLTGNSTNK
201	GNMAAIGAR	KWLESGASVG	VLYGHSRRGV	AQNYRVGGGG	QHIGNFGEEY
251	LERRKQQYFV	EGGLKFNAG	SGKWERDLQR	QYWKTKWYKK	YEDPQLQKY
301	IEEHDKSWRE	NLAPQYDITF	IDPSGLKQGS	AGNLFKLEYD	GVFNKYTAQF
351	RDNLNTRIGSR	KIINRNQYFN	YGLSLNPTYN	LNLTAAINSRG	RQKYPKGAKF
401	TGWGLLKDFE	TYNNAKILDL	NNATFRFLRP	ETELQTTLGF	NYFHNEYGRN
451	RFPEELGLFF	GDPDQDNGLY	SYLGRFGKDK	GLLPKSTIV	QFAGSGQFMT
501	FYFDALAKKD	LYRLNYSINA	INWRPGEIST	GVYGSENEK	RAFENSRAY
551	KEHCDPSGGL	YEVFLKYYKQ	KRANHVSVI	SADFDQYFMP	FAGYSRTHRM
601	FNIQENYFSQ	IGDSGVHTAL	KPERANTWQF	GENTYKGLL	KQDIDLGLKL
651	VCYRSRIDNY	IHNVYKQWMD	INGDIPSWVG	STGLAYTIRH	RNFKDVKVHK
701	GFELELNVDY	GRFTTNLSYA	YQKSTQPTNF	SDASEPNNA	SKEDQLKQGY
751	GLSRYSVALPR	DYGRLEVQTR	WLGKLTLLG	AMRYFGKSIR	ATAEERYIDG
801	TNGGNTSNVR	QLGKRSIKQT	ETLARQPLIF	DFYAAEYPPK	NLI FRAEVGN
851	LFDRRYIDPL	DAGNDAATQR	YYSFDPKDK	DEDVTGNADK	TLCNGKYGGT
901	SKSVLTNFR	GRTFLMTMSY	KF*		

ORF133ng-1 and ORF133-1 show 96.2% identity in 889 aa overlap:

	10	20	30	40	50	60
orf133ng-1.pep	SFRLKPICFYLMGVMLYHHSYAEADGRAGSEAIQVL	EDVHVAKKRV	PKDKKVF	TDA	RAV	
45	orf133-1			EAQIQVL	EDVHVAKKRV	PKDKKVF
				10	20	30
	70	80	90	100	110	120
orf133ng-1.pep	STRQDVFKSGENLNI	VRISIPGFTQQDKSSGIVSLN	IRGDSGFRVNTMVDGTTQT	FYS		
50	orf133-1	STRQDIFKSEN	LNIVRSIPGFTQQDKSSGIVSLN	IRGDSGFRVNTMVDGTTQT	FYS	
		40	50	60	70	80
	130	140	150	160	170	180
orf133ng-1.pep	TSTDAGRAGSSQFGASVDSNFIAGLD	VDVKGSGINSLAGSANLRTL	LGVDVVQGN			
55	orf133-1	TSTDAGRAGSSQFGASVDSNFIAGLD	VDVKGSGINSLAGSANLRTL	LGVDVVQGN		
		100	110	120	130	140
	190	200	210	220	230	240
orf133ng-1.pep	NTYGLLLKGLTGTNSTKGNMAAIGARKWLESGASVG	LVYGHRSRVAQNYRVGGGGQHI				
60	orf133-1	NTYGLLLKGLTGTNSTKGNMAAIGARKWLESGASVG	LVYGHRSRVAQNYRVGGGGQHI			
		160	170	180	190	200
	250	260	270	280	290	300
orf133ng-1.pep	GNFGEEYLERRKQQYFV	EGGLKFNAGSGKWERDLQRQYWKTKWYK	YEDPQLQKY	IEE		

10

15

20

25

30

35

40

45

50

55

65

70 In addition, ORF133ng-1 is homologous to a TonB-dependent receptor in *H.influenzae*:

spiP45114 YC17 HAEHN PROBABLE TONB-DEPENDENT RECEPTOR HI1217 PRECURSOR
 >g11075372|p|I|G64110 transferrin binding protein 1 precursor (tblp1) homolog -
 Haemophilus influenzae (strain Rd KW20) >g11574147 (U32801) transferrin binding
 protein 1 precursor (tblp1) [Haemophilus influenzae] Length = 913
 Score = 930 bits (2377), Expect = 0.0
 Identities = 476/921 (51%), Positives = 619/921 (66%), Gaps = 72/921 (7%)

Query: 38 QVLEDVHVAKVRPKDKKVFTRDAVSTRQDVFKSGENLDNIVRSIPGAFTQDDKSSGIV 97
 + L + V K + DKK FT+A+A STR++VFK + +D ++RSIPGAFTQDDK SG+V
 Sbjct: 29 ETLQQDVVEKVISNDKKFTEAKAKSTRENVFKETQTDIVQVRSIPGAFTQDDKSGGIV 88

Query: 98 SLNIRGDSGFGVRNIMVDGITQTFYSTSTADAGRAGGSSQFGVSDSNFIAGLDVVKGSFS 157
 S+NIRG++G GRVNIMVDG+TCTFYST+ D+G++GGSSQFGA++D NFIAG+DV K +FS
 Sbjct: 89 SVNIRGENGLGRVNIMVDGVTCTFYSTALDSQSGGSSQFGA++D NFIAGDVKNNSFS 148

Query: 158 GSAGINSLAGSANLRTLGVDVQVXXXXXXXXXXXXXXXXXXXXXAMAAIGARKWLESGA 217
 G++GIN+LAGSAN RTLG+DV+ M RKWL++G
 Sbjct: 149 GASGINALAGSANFRTLGVDVITDDKPFGIILKGMTSGNATSNFMTMAAGRKWLNDGG 208

Query: 218 SVGLYGHSSRRGVAQNYRVGGGGQHIGNFGEEYLERRKQOYFVQEGGLKFNAGSGKMERD 277
 VGV+YG+S+R V+Q+YR+ GGG+ + +G++ L + K+ YF + G N G+W D
 Sbjct: 209 YGVVVGYSQREVVSQDYRI--GGGERLASLGQDILAKEKAYF--RNAGYILNP--EQQWTPD 265

Query: 278 LQRQYWK-----TKWY-----KKYEDPQELQK---YIEE 303
 L +++W ++Y KK +D ++LQK IEE
 Sbjct: 266 LSKKHWSCKNPDYQKNGDCSYRIGSAARTREILQELLNGKKPKDIKQLQKNGDIEE 325

Query: 304 HDKSWRENLAPOYDITPIDPSGLKQOSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII 363
 DKS+ N QY + PI+P L+ +S +L K RKI AQ R L+ +IGSRKI
 Sbjct: 326 TDKSFERN--KDQYSVAPIEPGSLQSRSRSHLLKFEYGGDDHQNLAGQLRTLNDKIGSRKIE 384

Query: 364 NRNYQFNYGLSLNPTNLNLTAAYNSGRQYKPKAGFTGWLKDDFETYNNAKIDLNLNT 423
 NRNYQ NY + N Y +LNL RA++N G+ YKGG
 Sbjct: 385 NRNYQNYNYFNNNYSLDLNLMAAHNIGKTYIYKGGFFAGQVADKLITANVAVTIDNNS 444

Query: 424 AFTFLPRETELQTLTGNYFHNEYGNRFFPEELGLFDGPDQDNLGYSY--LGRFKGDGK 481
 TF LP+E +L+TTLGNYF NEY KNRFFPEEL LF++ D GLYS+ GR+ G K
 Sbjct: 445 HTFLLPKEIDLKTLTGNYFNEYGNRFFPEELSLFYNDASHDQGLYSHSKGRGYSGTKS 504

Query: 482 LLPQKSTIVQAGSQYENTFYFDAALKDIIYRLNYSNAINYRFGGEYTGNYGSENEFKR 541
 LLPQ+S I+Q+G Q F T YFD AL K IY LNYS N +Y F GEY GY
 Sbjct: 505 LLPQRSVILQPSGKQKFTYVFDALSKGIYHLNYSVNFTHYAFNGEYDV----- 555

Query: 542 AFGENSPAYKEHCDPSCGLYEPVLKKGKRRANNHVSISADFGDYFMFFAGYSRTHRMF 601
 EN+ + +EP+L K G K+A NHS ++SA+ DYFMFF YSRTHRMF
 Sbjct: 556 ---ENTAGQQ-----INPEILHKSGHKKAFNHSATLSAELSDYFMFFPTTYSRTHRMF 604

Query: 602 NIQEMYSQIGDSGVHTALKFERANTWQGFNTYKKGLKQDDILGLKILVGRYSRIDNYI 661
 NIQEM+FSQ+ ++GV+TALKFE+++T+Q GNTYKKGL QD+LG+KLVGRYSI N IYI
 Sbjct: 605 NIQEMFFSQSNAGVNTALKFEQSDTYQLGNTYKKGLFTQDDVLGKLVGRYSIFKNIY 664

Query: 662 HNVYGNKWDNLGDISPVGSGTGLAYTIRHRENFKKVHKHGFLELNDYGRFTNLSYAY 721
 HNVYGN ++ +P+W S G YTI H+N+K V K G ELE+ND GRFF N+SYAY
 Sbjct: 665 HNVYGVKW--RDMGTFWASNGFKYTIHAGNYKPIVKKSGVELEINYDMGRFFANVSYAY 722

Query: 722 QKSTQPTNFSDAESPNNASKEQDLKQGYGLSRVSALPRDYGRLEVGTRWLGKNTLGLGA 781
 Q++ QPTN++DAS PNNAS+ED LKQGYGLSRVS LP+DYGRLE+GTRW KLTLG A
 Sbjct: 723 QRTQPTNFDASPRPNNASQEDILKQGYGLSRVSMLPKDYGRLELGTWFDQKTLGLGA 782

Query: 782 MRYFGKSRIRATAERYIDGTNGGNTSNVRQLKGRSIKQETELARQPLIFDYAAYEPKKN 841
 RY+GKS RAT EE YI+G+ + +R+ ++K+TE + +Q+I D + +YEP K+
 Sbjct: 783 ARYFGSKRATIEEYINGSR--FKNTLRRENYAVKKTEDTKQPTIILDHVSYPEIKD 841

Query: 842 LIFRAEVKNLFDRRIIDPLDAGNDAAATQRYSSFPDKDKDEDVTNCNADKTLGNGKYGTS 901
 LI ++AEV+NL D+RY+DPLDAGNDAA+QRYSS + + C D + C GG+
 Sbjct: 842 LIIAEVEQNLDRKYVDPDLDAGNDAAQRYSS----NNSIECAQDSSAC-----GGSD 892

Query: 902 KSVLTNFAARGTFLMTMSYKF 922
 K+VL NFARGT+++++YKF
 Sbjct: 893 KTVLYNFAARGTYIILSNYKF 913

The underlined motif in the gonococcal protein (also present in the meningococcal protein) is predicted to be an ATP/GTP-binding site motif A (P-loop), and the analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 Example 104

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 885>

```

1 ATGAACCTGA TTTACGCTTA CATCATCCGT CAAATGGCGG TTATGGCGGT
51 TTACGCGCTC CTTGCTTTCC TCGCTTTGTA CAGCTTTTGT GAAATCCTGT
101 ACGAAACCGG CAACCTCGGC AAAGGACGTT ACGGCATATG GGAAATGCTG
151 GGCTACACCG CCCTCAAAAT GCCCGCCGCG GCCTACGAAC TGATTCCCTC
201 CGCGCTCCTT ATCGGCGGAC TGGTCTCCCT CAGCCAGCTT GCCCGCGCA
251 GCGAATCTAC CCGCATCAAA GGCACGGGCA TGAGACCAAA AAGCTGCTG
301 TTGATTCTGT CGCAGTTCCG TTTTATTTT GCTATTGCCA CGCTGCGCTG
351 CGGCGAATGG GTTGCGGCCA CACTGAGCCA AAAAGCCGAA AACATCAAA
401 CGCGCCGCAT CAACGGCAAA ATCAGCACCG GCAATACCGG CCTTTGGCTG
451 AAAGAAAAAA ACAGCGTGAT CAATGTGGCG GAAATGTGCG CGCACCAT..

```

This corresponds to the amino acid sequence <SEQ ID 886; ORF112>:

```

1 MNLISRYIIR QMAVMVAVYAL LAFLALYSFF EILYETGNLG KGSYGIWEML
51 GYTALKMPAR AYEILPLAVL IGGVLVLSQL AAGSELTVIK ASGMSTKKLL
101 LILSQFGFIF AIATVALGEW VAPTLSQKAE NIKAAAIN GK ISTGNTGLWL
151 KERNSVINVR EMLPDH...

```

Further work revealed further partial nucleotide sequence <SEQ ID 887>:

```

1 ATGAACCTGA TTTACGCTTA CATCATCCGT CAAATGGCGG TTATGGCGGT
51 TTACGCGCTC CTTGCTTTCC TCGCTTTGTA CAGCTTTTGT GAAATCCTGT
101 ACGAAACCGG CAACCTCGGC AAAGGACGTT ACGGCATATG GGAAATGCTG
151 GGCTACACCG CCCTCAAAAT GCCCGCCGCG GCCTACGAAC TGATTCCCTC
201 CGCGCTCCTT ATCGGCGGAC TGGTCTCCCT CAGCCAGCTT GCCCGCGCA
251 GCGAATCTAC CCGCATCAAA GGCACGGGCA TGAGACCAAA AAGCTGCTG
301 TTGATTCTGT CGCAGTTCCG TTTTATTTT GCTATTGCCA CGCTGCGCTG
351 CGGCGAATGG GTTGCGGCCA CACTGAGCCA AAAAGCCGAA AACATCAAA
401 CGCGCCGCAT CAACGGCAAA ATCAGCACCG GCAATACCGG CCTTTGGCTG
451 AAAGAAAAAA ACAGCTKAT CAATGTGGCG GAAATGTGCG CCGACCATAC
501 GCTTTTGGGC ATCAAAATTT GGGCGCGCAA CGATAAAAC GAATTGGCAG
551 AGGCGAGTGA AGCGGATTCC GCGCTTTTGA ACAGCGACCG CAGTTGGCAG
601 TTGAAAAACA TCCGCGCGCAG CAGCGTTTGC GAAGACAAG TCGAGGTCTC
651 TATTGGCGCT GAAGAAAAC TCGCGGATTC CGTCAACACG AACCTGATGG
701 ACGTATTGCT GCTCAAAACC GACCAATGT CGGTCGGGCA ACTGACACCC
751 TACATCCGCC ACCTCCAAAA CAACAGCCAA AACACCGGAA TCTAGCCCAT
801 CGCATGGTGG CGCAAAATGG TTTACCCCGC CGCAGCTGG GTGATGGCGC
851 TCGTGCCTTT TGCCTTTACC CCGCAACCA CCGCGCACCG CAATATGGCG
901 TTAATACTCT TCGCGCGCAT CTGTGTGGGA TTGCTGTTC ACCTTGGCGG
951 ACGGCTCTTT GGGTTTACCA GCCAATCGG...

```

This corresponds to the amino acid sequence <SEQ ID 888; ORF112-1>:

```

1 MNLISRYIIR QMAVMVAVYAL LAFLALYSFF EILYETGNLG KGSYGIWEML
51 GYTALKMPAR AYEILPLAVL IGGVLVLSQL AAGSELTVIK ASGMSTKKLL
101 LILSQFGFIF AIATVALGEW VAPTLSQKAE NIKAAAIN GK ISTGNTGLWL
151 KERNSVINVR EMLPDHLLG IKIWARNDK ELAAVEADPS AVINSDGSWQ
201 LKNIRRLTLG EDKVEYSIAA EENWPISVKR NMDVLLVPR DQMSVELTNG
251 YIRHLQNSQ NTRIYAIWW RKLVPAAAW VMALVAFAT PQTTRHNGMG
301 LKLPGGICXG LLFPLAGRLF GFTSQL...

```

Computer analysis of this amino acid sequence predicts two transmembrane domains and gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF112 shows 96.4% identity over a 166aa overlap with an ORF (ORF112a) from strain A of *N.*

meningitidis:

		10	20	30	40	50	60
5	orf112.pep	MNLI	SR	YI	IR	QM	AV
	orf112a	MNLI	SR	YI	IR	QM	AV
		10	20	30	40	50	60
10	orf112.pep	AYEL	IP	LA	VL	IG	GL
	orf112a	AYEL	IP	LA	VL	IG	GL
		70	80	90	100	110	120
15	orf112.pep	VAPT	LS	QK	AE	NI	KA
	orf112a	VAPT	LS	QK	AE	NI	KA
		130	140	150	160		
20	orf112.pep	ELAE	AV	EA	DS	AV	LN
	orf112a	ELAE	AV	EA	DS	AV	LN
		190	200	210	220	230	240

The ORF112a nucleotide sequence <SEQ ID 889> is:

25	1	ATGAACCTGA	TTTCAAGTTA	CATCATCCGT	CAAAATGGCGG	TTATGGCGGT
	51	TTACGGCGCTC	CTTGCCCTTCC	TGCGTTTGTG	CAGCTTTTTT	GAATACCTGT
	101	ACGAACCGCT	CAACCTCGCG	AAAGCGAGTT	ACGGCATATG	GGAAATGNTG
	151	GGNTACACCG	CCCTCAAAAT	GNCCGCCCGC	GGCTACGAC	TGATCCCGCT
	201	CGCCCTCTCT	ATCGCGCGAC	TGCTCTCTNT	CAGCCAGCTT	CGCGCGCGCA
30	251	CGCAACTGAN	CGTCATCAA	CGCAGCGGCA	TGAGACCCAA	AAAGCTCTGT
	301	TTGATTCTGT	CGCAGTTCGG	TTTTATTTTT	GCTATTGCCA	CGCTCGCGCT
	351	CGCGGAATGG	GTTGCGGCCA	CATGAGCCCA	AAAAGCGGAA	AACATCAAAG
	401	CGCGGCCCAT	CAACGGCRAA	ATCAGTACCG	GCAATACCGG	CCTTTGGCTG
	451	AAAGAAAAAA	ACAGCATTAT	CAATGTGCGC	GAAATGTTGC	CGACCATAC
35	501	CCTGCTGGGC	ATTAAATCT	GGGCGCGCAA	CGATAAAAC	GAATGGCAG
	551	AGGCAGTGG	AGCCGATTCC	CGCGTTTGA	ACAGCGACGG	CAGTTGGCAG
	601	TTGAAAAACA	TCCGCGCGAG	CACGCTTGGC	GAAGACAAAG	TGAGGCTCTC
	651	TATTGCGGCT	GAAGAAAANT	GGCGGATTTC	CGTCAACCG	AACCTGATGG
40	701	ACGTATTGCT	CGTCAACCC	GACCAATGT	CGCTCGGCGA	ACTGACCACC
	751	TACATCGGCC	ACCTCCAAAN	NNACAGCCAA	AACACCGGAA	TCTAGCCCAT
	801	CGCATGTTGG	CGCAAAATGG	TTTACCCCGC	CGCAGCCTGG	GTGATGGCGC
	851	TCGTGCGCTT	TGCTTTTACC	CCGCAAAACA	CCGCGCACGG	CAATATGGGG
	901	TTAAANATCT	TCGCGCGCAT	CTGTCTCGGA	TTGCTGTTC	ACCTTGGCGG
	951	NCGGCTCTTC	NGGTTTACCA	GCCAACTCTA	CGCGATCCCG	CCCTTCTCG
45	1001	NCGGGCGACT	ACCTACCAT	GCTTCCGCT	TGCTCGCGCT	TTGGCTGATA
	1051	CGCAACAGG	AAAAAGCTA	A		

This encodes a protein having the amino acid sequence <SEQ ID 890>:

	1	MNLSIRYIIR	QMAMVAVYAL	LAFALYXSSP	EILVETGNLG	KGSGYIWEKK
	51	GYTLKXKXAR	AYEMLFAYL	IGGLVEXSFL	ANGELXVL	ASVMS2EKK
50	101	LILSOPR	YALFALGEN	VAPFLSQAK	NITKAARENG	ITSGTNGNGL
	151	KEKNISINVR	EMLPDHTLGL	IKIWARNDKN	ELAEAAEAS	AGN2SDGSLG
	201	LKNIRRSTLG	EKKVEVSI	EEXWPISSVR	NLMDVLVLC	DOMSVELGTU
	251	YIRHQQXSQ	NTRYIATAWM	RKLVIYFAAAW	NLMVAFAFT	POTTRHGNMG
	301	LXKFGGICL	LLFHLGAGRL	XFTSOLYGTP	PFLXGLPTI	AFALLAVNLI
55	351	RKOEKR*				

ORF112a and ORF112-1 show 96.3% identity in 326 aa overlap:

orf112a.pep	MNLI	SR	YI	IR	QM	AV	LA	FL	AL	Y	FF	E	I	Y	ET	GN	L	K	G	S	Y	GI	WE	MX	GY	T	L	A	K	X	A	R										
orf112-1	MNLI	SR	YI	IR	QM	AV	LA	FL	AL	Y	FF	E	I	Y	ET	GN	L	K	G	S	Y	GI	WE	MX	GY	T	L	A	K	X	A	R										
orf112a.pep	AYEL	IP	LA	VL	IG	GL	VS	X	S	Q	L	A	AG	S	EL	X	VI	K	AS	GM	ST	K	L	L	L	L	I	L	S	Q	F	I	A	I	A	T	A	L	V	AL	GE	W

10

15

25

30

35

50

60

1	MNLISYIIR	QMAVMAYVAL	LAFLALYSFF	EILYETGNLG	KGSYGIWEML
51	GYTKALPBR	AYELMALPVL	IGGLASLSPQ	ANGSELAVIK	ASQMSTKLLL
101	LILSQEGPFI	AVLAALVGEV	TPATLSQAKE	NILKAATINGK	ISTGNTGLML
151	KEKTSINVR	GMLPDHTLIG	KITWARDNK	EALAEADVDS	AVLNSDGSWO
201	LKNIRRSIR	TDKLTLSAAA	ETWPIAVRR	NIMDLVLVPP	QMSVELITTV
251	YRIHLQNNQS	NTQIYALAKW	RKLVPYPAW	VMALVAFPT	PQTRRHNGN
301	LKLPGCTIG	LLEHILAGLF	GKTSLOYSPL	FFALGAPVIT	AFALLAVWLI

351 RKQEK*
*

ORF112ng and ORF112-1 show 94.2% identity in 326 aa overlap:

		10	20	30	40	50	60
5	orf112ng	MNLI	SRYYIRQM	MAVM	AVYALLA	FLALYS	FPEILYETGNLKGSGYIWEMLGYTALKMPAR
	orf112-1	MNLI	SRYYIRQM	MAVM	AVYALLA	FLALYS	FPEILYETGNLKGSGYIWEMLGYTALKMPAR
		10	20	30	40	50	60
		70	80	90	100	110	120
10	orf112ng	AYEL	IMPLAVLIGGLASLSQ	LAAGSEL	AVIKASGMSTK	KKLLIL	LSQFGFIPAIATVALGEM
	orf112-1	AYEL	IMPLAVLIGGLVLSQ	LAAGSEL	TVIKASGMSTK	KKLLIL	LSQFGFIPAIATVALGEM
		70	80	90	100	110	120
15	orf112ng	VAPT	LSQKAENIKAAA	INGKISTG	NGLWKEKTS	IIINVRG	MLPDHTLLGIKIWARNDKN
	orf112-1	VAPT	LSQKAENIKAAA	INGKISTG	NGLWKEKTS	IIINVRG	MLPDHTLLGIKIWARNDKN
		130	140	150	160	170	180
20	orf112ng	ELAE	AVEADSAVLNS	DGSQLKNIR	RSINGTDKI	ETSAAEET	WPIAVRRNLM
	orf112-1	ELAE	AVEADSAVLNS	DGSQLKNIR	RSINGTDKI	ETSAAEET	WPIAVRRNLM
		190	200	210	220	230	240
25	orf112ng	ELAE	AVEADSAVLNS	DGSQLKNIR	RSINGTDKI	ETSAAEET	WPIAVRRNLM
	orf112-1	ELAE	AVEADSAVLNS	DGSQLKNIR	RSINGTDKI	ETSAAEET	WPIAVRRNLM
		190	200	210	220	230	240
		250	260	270	280	290	300
30	orf112ng	DOMS	VGELTTYIRHLQ	NNSONTQI	YAIWWRKLV	YPAAAW	VMALVAFETPQTTRHGNMG
	orf112-1	DOMS	VGELTTYIRHLQ	NNSONTQI	YAIWWRKLV	YPAAAW	VMALVAFETPQTTRHGNMG
		250	260	270	280	290	300
35	orf112ng	LKLF	GGICLGLLPH	LHAGRLFG	PTSQLYGT	PPFLAGAL	PTIAPALLAVWLIRKQEKRX
	orf112-1	LKLF	GGICLGLLPH	LHAGRLFG	PTSQLYGT	PPFLAGAL	PTIAPALLAVWLIRKQEKRX
		310	320	330	340	350	

This analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their
 40 epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

It will be appreciated that the invention has been described by means of example only, and that
 modifications may be made whilst remaining within the spirit and scope of the invention.

TABLE I – PCR primers

ORF	Primer	Sequence	Restriction sites
ORF 1	Forward	CGCGGATCCGCTAGC-GGACACACTTATTTCGG	BamHI-NheI XhoI
	Reverse	CCCGCTCGAG-CCAGCGGTAGCCTAATT	
ORF 2	Forward	GCGGATCCCATATG-TTTGATTTCGGTTTGGG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-GACGGCATAACGGCG	
ORF 2-1	Forward	GCGGATCCCATATG-TTTGATTTCGGTTTGGG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TGATTACGGACGCGCA	
ORF 4	Forward	GCGGATCCCATATG-TGCGGAGGTCAAAAAGAC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TTTGGCTGCGCCTTC	
ORF 5	Forward	GGAATTCCATATGGCCATGG-TGGAAGGCGCACAAAC	NdeI-NcoI BamHI XhoI
	Reverse	CCCGCTCGAG-GACTGTGCAAAAACGG	
ORF 6	Forward	CGCGGATCCCATATG-ACCCGTCAATCTCTGCA	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TGCGCCGAACACTTTC	
ORF 7	Forward	CGCGGATCCGCTAGC-GCGCTGCTTTTTGTTC	BamHI-NheI XhoI
	Reverse	CCCGCTCGAG-TTTCAAAATATATTTCGGA	
ORF 8	Forward	GCGGATCCCATATG-GCTCAACTGCTTCGTAC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-AGCAGGCTTTGGCGC	
ORF 9	Forward	CGCGGATCCCATATG-CCGAAGGAAGTCGGAAA	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TTTCGAGGTTTTTCGGG	
ORF 10	Forward	GCGGATCCCATATG-GACACAAAAGAAATCCTC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TAATGGGAACCTTGTTTT	
ORF 11	Forward	GCGGATCCCATATG-GCGGTCAACCTCTACG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-GGAAACGACTTCGCC	
ORF 13	Forward	CGCGGATCCCATATG-GCTCTGCTTTCCGCGC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-AGGTTGTGTGATAATAAG	
ORF 15	Forward	GGAATTCCATATGGCCATGG-GCGGGACACTGACAG	NdeI-NcoI BamHI XhoI
	Reverse	CCCGCTCGAG-AGGTTGGCCTTGCTATG	
ORF 17	Forward	GGAATTCCATATGGCCATGG -TTGCCGCGCTGTTTCG	NdeI-NcoI

	Forward Reverse	CGGGATCC-ATTGCCGGCCTGTTCC CCCGCTCGAG-AAGCAGGTTGTACAGC	BamHI XhoI
ORF 18	Forward Reverse	GCGGATCCCATATG-ATTTTCTGCATTTGGAT CCCGCTCGAG-TCTTCCAATTTCTGAAAGC	BamHI-NdeI XhoI
ORF 19	Forward Forward Reverse	GGAATTCATATGGCCATGG -TCGCCAGTGTTTTACC CGGGATCC-TTCGCCAGTGTTTACC CCCGCTCGAG-GGTGTTTTGAAGCTGCC	NdeI-NcoI BamHI XhoI
ORF 20	Forward Forward Reverse	GGAATTCATATGGCCATGG -TCGGCGCGGGTATG CGGGATCC-TTCGGCGCGGGTATG CCCGCTCGAG-CGGCGAGCGAGAGCA	NdeI-NcoI BamHI XhoI
ORF 22	Forward Forward Reverse	GGAATTCATATGGCCATGG-TGATTAAAAACAAAAAGGTCT CGGGATCC-ATGATTAAAAACAAAAAGGTCTAAACC CCCGCTCGAG-ATTATGATAGCGGCC	NdeI-NcoI BamHI XhoI
ORF 23	Forward Reverse	CGCGGATCCCATATG-GATGTTTCTGTTTCAGAC CCCGCTCGAG-TTAAACCGATAGGTAACG	BamHI-NdeI XhoI
ORF 24	Forward Forward Reverse	GGAATTCATATGGCCAIGG -TGATGCCGAAATGGTG CGGGATCC-ATGATGCCGGAATGGTG CCCGCTCGAG-TGTGAGCTGGCGCA	NdeI-NcoI BamHI XhoI
ORF 25	Forward Reverse	GCGGATCCCATATG-TATCGCAAACGATTGC CCCGCTCGAG-ATCGATGGAATAGCCG	BamHI-NdeI XhoI
ORF 26	Forward Reverse	GCGGATCCCATATG-CAGCTGATCGACTATTC CCCGCTCGAG-GACATCGGCGGTTTT	BamHI-NdeI XhoI
ORF 27	Forward Forward Reverse	GGAATTCATATGGCCATGG-AGACCTATTCTGTTTA CGGGATCC- CAGACCTATTCTGTTTATTTAATC CCCGCTCGAG-GGGTTCGATTAAATAACCAT	NdeI-NcoI BamHI XhoI
ORF 28	Forward Forward Reverse	GGAATTCATATGGCCATGG-ACGGCTGTACGTTGATGT CGGGATCC-AACGGCTGTACGTTGATG CCCGCTCGAG-TTGTCTAGAGGAATTCGCG	NdeI-NcoI BamHI XhoI
ORF 29	Forward Forward Reverse	GCGGATCCCATATG -AACGGTTTGGATGCCCC CGCGGATCCGCTAGC-AACGGTTTGGATGCCCC CCCGCTCGAG-TTGTCTAAGTTCTGATATG	BamHI-NdeI BamHI-NheI XhoI
ORF 32	Forward Reverse	CGCGGATCCCATATG-AATACTCCTCCTTTTG CCCGCTCGAG-GCGTATTTTTGATGCTTTG	BamHI-NdeI XhoI
ORF 33	Forward Reverse	GCGGATCCCATATG -ATTGATAGGGATCGTATG CCCGCTCGAG-TTGATCTTTCAAACGGCC	BamHI-NdeI XhoI

ORF 35	Forward Forward Reverse	GCGGATCCCATATG-TTCAGAGCTCAGCTT CGCGATCCGCTAGC-TTCAGAGCTCAGCTT CCCGCTCGAG-AAACAGCCATTGAGCGA	BamHI-NdeI BamHI-NheI XhoI
ORF 37	Forward Reverse	GCGGATCCCATATG-GATGACGTATCGGATTTT CCCGCTCGAG-ATAGCCCGCTTTCAGG	BamHI-NdeI XhoI
ORF 58	Forward Reverse	CGCGGATCCGCTAGC-TCCGAACGCGAGTGGAT CCCGCTCGAG-AGCATTGTCCAAGGGGAC	BamHI-NheI XhoI
ORF 65	Forward Forward Reverse	GGAATTCCATATGGCCATGG -TGCTGTATCTGAATCAAG CGGGATCC-TTGCTGTATCTGAATCAAGG CCCGCTCGAG-CCGCATCGGCAGACA	 NdeI-NcoI BamHI XhoI
ORF 66	Forward Reverse	GCGGATCCCATATG-TACGCATTACCGCGG CCCGCTCGAG-TGGATTTTGCAGAGATGG	BamHI-NdeI XhoI
ORF 72	Forward Reverse	CGCGGATCCCATATG- AATGCAGTAAAAATATCTGA CCCGCTCGAG-GCCTGAGACCTTTGCAA	BamHI-NdeI XhoI
ORF 73	Forward Reverse	GCGGATCCCATATG-AGATTTTTCGGTATCGG CCCGCTCGAG-TTCATCTTTTTCATGTTG	BamHI-NdeI XhoI
ORF 75	Forward Reverse	GCGGATCCCATATG- TCTGTCTTTCAAACGGC CCCGCTCGAG-TTTGTTTTGCAAGACAG	BamHI-NdeI XhoI
ORF 76	Forward Reverse	GATCAGCTAGCCATATG-AAACAGAAAAAACCGC CGGGATCC-TTACGGTTTGACACCGTT	NheI-NdeI BamHI
ORF 79	Forward Reverse	CGCGGATCCCATATG-GTTTCGCGCGCGG CCCGCTCGAG-GTGTGATGCGCTTCG	BamHI-NdeI XhoI
ORF 83	Forward Reverse	GCGGATCCCATATG-AAAACCTGTCTGCTGC CCCGCTCGAG-GCCGCCTTTGCGGC	BamHI-NdeI XhoI
ORF 84	Forward Reverse	GCGGATCCCATATG-GCAGAGATCTGTTTG CCCGCTCGAG-GTTTGCCGATCCGACCA	BamHI-NdeI XhoI
ORF 85	Forward Reverse	CGCGGATCCCATATG- GCGGTTTGGGCGGA CCCGCTCGAG-TCGGCGCGCGGGC	BamHI-NdeI XhoI
ORF 89	Forward Forward Reverse	GGAATTCCATATGGCCATGG-CCATACCTTCTTATCA CGGGATCC-GCCATACCTTCTTATCAGAG CCCGCTCGAG-TTTTTTGCGATTAGAAAAAGC	NdeI-NcoI BamHI XhoI
ORF 97	Forward	GCGGATCCCATATG-CATCTGCCAGCGAAC	BamHI-NdeI

	Reverse	CCCGCTCGAG-TTCGCCTACGGTTTTTTG	XhoI
ORF 98	Forward	GCGGATCCCATATG-ACGTAACCTGCGG	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTGTGTTCGGCAAATC	XhoI
ORF 100	Forward	GCGGATCCCATATG-TCGGGCTTTACACCG	BamHI-NdeI
	Reverse	CCCGCTCGAG-ACGGGTTCGGCGGAA	XhoI
ORF 101	Forward	GCGGATCCCATATG-ATTTATCAAGAAACCTC	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTTCGCCTTTCATGT	XhoI
ORF 102	Forward	GCGGATCCCATATG-GCAGGGCTGTTTACC	BamHI-NdeI
	Reverse	CCCGCTCGAG-AAACGGTTTGAACACGAC	XhoI
ORF 103	Forward	GCGGATCCCATATG-AACCACGACATCAC	BamHI-NdeI
	Reverse	CCCGCTCGAG-CAGCCACAGGACGGC	XhoI
ORF 104	Forward	GCGGATCCCATATG-ACGTGGGGAACGC	BamHI-NdeI
	Reverse	CCCGCTCGAG-GCGGCGTTTGAACGGC	XhoI
ORF 105	Forward	GCGGATCCCATATG-ACCAATTTCAAACCCCTC	BamHI-NdeI
	Reverse	CCCGCTCGAG-TAAACGAATGCCGTCCAG	XhoI
ORF 106	Forward	GCGGATCCCATATG-AGGATAACCGACGGCG	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTGTTCCGATGATGTT	XhoI
ORF 109	Forward	GCGGATCCCATATG-GAAGATTTATATATAATACTCG	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATCAGCTTCGAACCGAAG	XhoI
ORF110	Forward	AAAGAATTC-ATGAGTAAATCCCGTAGATCTCCC	EcoRI
	Reverse	AAACTGCAG-GGAAAACCATCCGCACTCTGCC	PstI
ORF111	Forward	AAAGAATTC-GCACCGCAAAAGCAAAAACCGCA	EcoRI
	Reverse	AAACTGCAG-TCTGCGCGT TTTGGGCAGGGTGG	PstI
ORF113	Forward	AAAGAATTC-ATGAACAAAACCTCTATCGTGTATTTCAACCG	EcoRI
	Reverse	AAACTGCAG-TTACGAATGCCTGCTTCTCGACCGTACTG	PstI
ORF115	Forward	AAAGAATTC-TTGCTTGTGCAACAGAAAAGACGG	EcoRI
	Reverse	AAAAAGTCGAC-CTATTTTTAGGGGCTTTTGTGTTGAAAAGCCTGCC	SalI
ORF119	Forward	AAAGAATTC-TACAACATGTATCAGGAAAACCAATACCG	EcoRI
	Reverse	AAACTGCAG-TTATGAAAACAGGCGCAGGGCGGTTTTGCC	PstI
ORF120	Forward	AAAGAAATTC-GCAAGGCTACCCCAATCCGCCGTG	EcoRI
	Reverse	AAACTGCAG-CGGTTTGGCTGCCTGGCGTGTAT	PstI
ORF121	Forward	AAAGAATTC-GCCTTGGTCTGGCTGGTTTTCGC	EcoRI
	Reverse	AAACTGCAG-TCATCCGCCACCCACCTCGGCATCCATC	PstI

ORF122	Forward Reverse	AAAAAAGTCGAC-ATGTC TTACCGGCAAGCAGTTC TCC AAACTGCGAG-TCAGGAACACAAACGATGACGAATATCCGTATC	SalI PstI
ORF125	Forward Reverse	AAAGAATTC-GCGCTGTTTTTTGCGGCGGCGTAT AAACTGCGAG-CGCCGTTTCAAGACGAAAAAGTCG	EcoRI PstI
ORF126	Forward Reverse	AAAGAATTC-GCGGAAACGGTCGAAG AAACTGCGAG-TTAATCTTGCTCCGATATAC	EcoRI PstI
ORF127	Forward Reverse	AAAGAATTC-ATGACTGATAATCGGGGTTTACG AAAAAAGTCGAG-CTTAAGTAAC TTGACGTCTTATC	EcoRI SalI
ORF128	Forward Reverse	AAAGAATTC-ATGCAAGCTGTCGCTACAGGCC AAACTGCGAG-CTATTGCAATGCGCGCCGGGGAATGTTGACAGGGG	EcoRI PstI
ORF129	Forward Reverse	AAAGAATTC-ATGGATTTTCGTTTGACATTATTTACGAATACCG AAACTGCGAG-TTATTTTTGATGAAATTTTGGGCGG	EcoRI PstI
ORF130	Forward Reverse	AAAGAATTC-GCAGTACTTGCCATCTCGGTGCG AAACTGCGAG-CTCCGGATCGTCTGAACGCATT	EcoRI PstI
ORF 131	Forward Reverse	GCGGATCCCATATG-GAAATTCGGGCAATAAAAT CCCGCTCGAG-CCAGCGGACGCGTTC	BamHI-NdeI XhoI
ORF 132	Forward Reverse	GCGGATCCCATATG-AAAGAAGCGGGGTTTG CCCGCTCGAG-CCAATCTGCCAGCCGT	BamHI-NdeI XhoI
ORF 133	Forward Reverse	CGCGGATCCCATATG-GAAGATGCAGGGCGCG CCCGCTCGAG-AAACTTG TAGTCATCGT	BamHI-NdeI XhoI
ORF 134	Forward Reverse	GCGGATCCCATATG-TCTGTGCAAGCAGTATTG CCCGCTCGAG-ATCCTGTGCCAATGCG	BamHI-NdeI XhoI
ORF 135	Forward Reverse	GCGGATCCCATATG-CCGTCTGAAAAAGCTTT CCCGCTCGAG-AAATACCGCTGAGGATG	BamHI-NdeI XhoI
ORF 136	Forward Reverse	CGCGGATCCGCTAGC-ATGAAGCGCGTATAGCC CCCGCTCGAG-TTCCGAATATTGGAACTTTT	BamHI-NheI XhoI
ORF 137	Forward Reverse	CGCGGATCCCATATG-GGCACGGCGGAAATA CCCGCTCGAG-ATAACGGTATGCCGCC	BamHI-NdeI XhoI
ORF 138	Forward Reverse	GCGGATCCCATATG-TTTCGTTTACAATTCAGGC CCCGCTCGAG-CGGCGTTTTATAGCGG	BamHI-NdeI XhoI
ORF 139	Forward Reverse	GCGGATCCCATATG-GCTTTTTTGCGGTAATG CCCGCTCGAG-TAACGTTTCCGTGCGTTT	BamHI-NdeI XhoI

ORF 140	Forward Reverse	GCGGATCCCATATG-TTGCCCACAGGCAGC CCCGCTCGAG-GACGATGGCAACAGC	BamHI-NdeI XhoI
ORF 141	Forward Reverse	GCGGATCCCATATG-CCGTCTGAAGCAGTCT CCCGCTCGAG-ATCTGTTGTTTTAAATATT	BamHI-NdeI XhoI
ORF 142	Forward Reverse	GCGGATCCCATATG-GATAATTCTGGTAGTGAAG CCCGCTCGAG-AAACGTATAGCCTACCT	BamHI-NdeI XhoI
ORF 143	Forward Reverse	GCGGATCCCATATG-GATACCGCTTTGAACCT CCCGCTCGAG-AATGGCTTCCGCAATATG	BamHI-NdeI XhoI
ORF 144	Forward Reverse	GCGGATCCCATATG-ACCTTTTACAACGTTTGC CCCGCTCGAG-AGATTGTTGTTGTTTTTCG	BamHI-NdeI XhoI
ORF 147	Forward Reverse	GCGGATCCCATATG-TCTGTCTTTCAAACGGC CCCGCTCGAG-TTTGTTTTGCAAGACAG	BamHI-NdeI XhoI

NB:

- restriction sites are underlined
- for ORFs 110-130, where the ORF itself carries an *EcoRI* site (eg. ORF122), a *SaI* site was used in the forward primer instead. Similarly, where the ORF carries a *PstI* site (eg. ORFs 115 and 127), a *SaI* site was used in the reverse primer.

TABLE II – Summary of cloning, expression and purification

ORF	PCR/cloning	His-fusion expression	GST-fusion expression	Purification
orf 1	+	+	+	His-fusion
orf 2	+	+	+	GST-fusion
orf 2.1	+	n.d.	+	GST-fusion
orf 4	+	+	+	His-fusion
orf 5	+	n.d.	+	GST-fusion
orf 6	+	+	+	GST-fusion
orf 7	+	+	+	GST-fusion
orf 8	+	n.d.	n.d.	
orf 9	+	+	+	GST-fusion
orf 10	+	n.d.	n.d.	
orf 11	+	n.d.	n.d.	
orf 13	+	n.d.	+	GST-fusion
orf 15	+	+	+	GST-fusion
orf 17	+	n.d.	n.d.	
orf 18	+	n.d.	n.d.	
orf 19	+	n.d.	n.d.	
orf 20	+	n.d.	n.d.	
orf 22	+	+	+	GST-fusion
orf 23	+	+	+	His-fusion
orf 24	+	n.d.	n.d.	
orf 25	+	+	+	His-fusion
orf 26	+	n.d.	n.d.	
orf 27	+	+	+	GST-fusion
orf 28	+	+	+	GST-fusion
orf 29	+	n.d.	n.d.	
orf 32	+	+	+	His-fusion
orf 33	+	n.d.	n.d.	
orf 35	+	n.d.	n.d.	
orf 37	+	+	+	GST-fusion
orf 58	+	n.d.	n.d.	
orf 65	+	n.d.	n.d.	
orf 66	+	n.d.	n.d.	
orf 72	+	+	n.d.	His-fusion
orf 73	+	n.d.	+	n.d.
orf 75	+	n.d.	n.d.	
orf 76	+	+	n.d.	His-fusion
orf 79	+	+	n.d.	His-fusion
orf 83	+	n.d.	+	n.d.
orf 84	+	n.d.	n.d.	

orf 85	+	n.d.	+	GST-fusion
orf 89	+	n.d.	+	GST-fusion
orf 97	+	+	+	GST-fusion
orf 98	+	n.d.	n.d.	
orf 100	+	n.d.	n.d.	
orf 101	+	n.d.	n.d.	
orf 102	+	n.d.	n.d.	
orf 103	+	n.d.	n.d.	
orf 104	+	n.d.	n.d.	
orf 105	+	n.d.	n.d.	
orf 106	+	+	+	His-fusion
orf 109	+	n.d.	n.d.	
orf 110	+	n.d.	n.d.	
orf 111	+	+	n.d.	His-fusion
orf 113	+	+	n.d.	His-fusion
orf 115	n.d.	n.d.	n.d.	
orf 119	+	+	n.d.	His-fusion
orf 120	+	+	n.d.	His-fusion
orf 121	+	n.d.	n.d.	
orf 122	+	+	n.d.	His-fusion
orf 125	+	+	n.d.	His-fusion
orf 126	+	+	n.d.	His-fusion
orf 127	+	+	n.d.	His-fusion
orf 128	+	n.d.	n.d.	
orf 129	+	+	n.d.	His-fusion
orf 130	+	n.d.	n.d.	
orf 131	+	+	+	n.d.
orf 132	+	+	+	His-fusion
orf 133	+	n.d.	+	GST-fusion
orf 134	+	n.d.	n.d.	
orf 135	+	n.d.	n.d.	
orf 136	+	n.d.	n.d.	
orf 137	+	n.d.	+	GST-fusion
orf 138	+	n.d.	+	GST-fusion
orf 139	+	n.d.	n.d.	
orf 140	+	n.d.	n.d.	
orf 141	+	n.d.	n.d.	
orf 142	+	n.d.	n.d.	
orf 143	+	n.d.	n.d.	
orf 144	+	n.d.	+	n.d.
orf 147	+	n.d.	n.d.	

CLAIMS

1. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, and 8.

2. A nucleic acid molecule which encodes a protein according to claim 1.

3. A nucleic acid molecule according to claim 2, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, and 7.

4. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, & 892..

5. A protein having 50% or greater sequence identity to a protein according to claim 4.

6. A protein comprising a fragment of an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, & 892..
7. An antibody which binds to a protein according to any one of claims 4 to 6.
8. A nucleic acid molecule which encodes a protein according to any one of claims 4 to 6.
9. A nucleic acid molecule according to claim 8, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289,

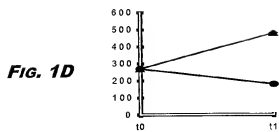
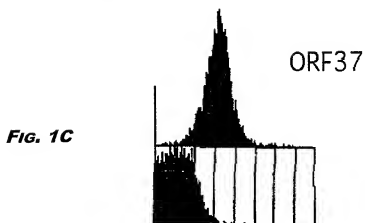
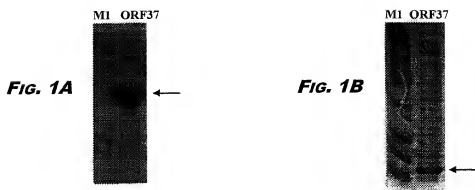
291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329,
 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369,
 371, 373, 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409,
 411, 413, 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449,
 5 451, 453, 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489,
 491, 493, 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529,
 531, 533, 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569,
 571, 573, 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609,
 611, 613, 615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649,
 10 651, 653, 655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689,
 691, 693, 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729,
 731, 733, 735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769,
 771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809,
 811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849,
 15 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889,
 & 891..

10. A nucleic acid molecule comprising a fragment of a nucleotide sequence selected from the
 group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39,
 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91,
 20 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133,
 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173,
 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213,
 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253,
 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293,
 25 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333,
 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373,
 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409, 411, 413,
 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449, 451, 453,
 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489, 491, 493,
 30 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529, 531, 533,
 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569, 571, 573,
 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609, 611, 613,
 615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649, 651, 653,
 655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689, 691, 693,
 35 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729, 731, 733,

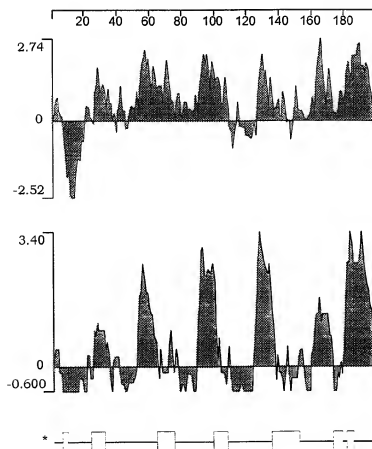
735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769, 771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809, 811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849, 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889, & 891..

- 5 11. A nucleic acid molecule comprising a nucleotide sequence complementary to a nucleic acid molecule according to any one of claims 8 to 10.
12. A nucleic acid molecule comprising a nucleotide sequences having 50% or greater sequence identity to a nucleic acid molecule according to any one of claims 8-11.
13. A nucleic acid molecule which can hybridise to a nucleic acid molecule according to any one of claims 8-12 under high stringency conditions.
- 10 14. A composition comprising a protein, a nucleic acid molecule, or an antibody according to any preceding claim.
15. A composition according to claim 14 being a vaccine composition or a diagnostic composition.
- 15 16. A composition according to claim 14 or claim 15 for use as a pharmaceutical.
17. The use of a composition according to claim 14 in the manufacture of a medicament for the treatment or prevention of infection due to Neisserial bacteria.

1/24

FIGURE 1

2/24

Fig 1E

3/24

FIGURE 2

Fig. 2A

M1 ORF5

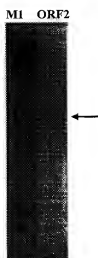
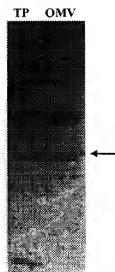
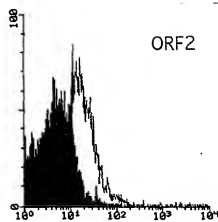


Fig. 2B

TP



4/24

FIGURE 3**Fig. 3A****Fig. 3B****Fig. 3C****Fig. 3D**

5/24

FIGURE 4

Fig. 4A

M1 ORF15

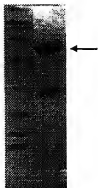


Fig. 4B

M2 ORF15

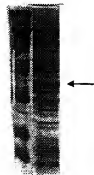
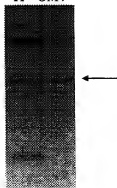
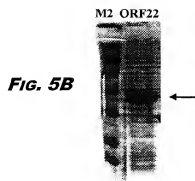
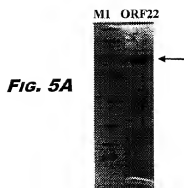
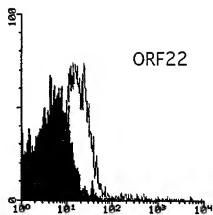


Fig 4C

TP OMV



6/24

FIGURE 5**FIG. 5C**

7/24

FIGURE 6

FIG. 6A

M1 ORF28

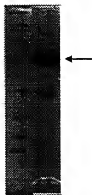


FIG. 6B

M2 ORF28

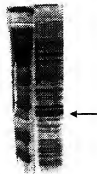


FIGURE 7

FIG. 7A

M1 ORF32

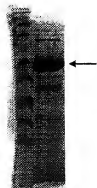
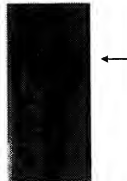
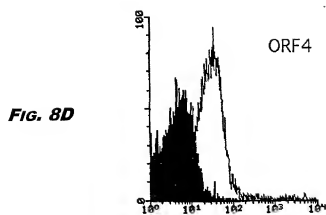
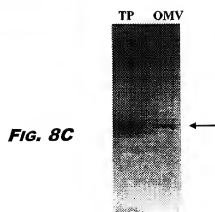
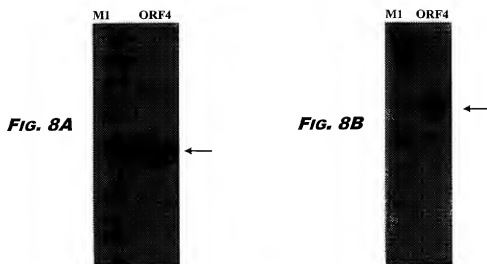


FIG. 7B

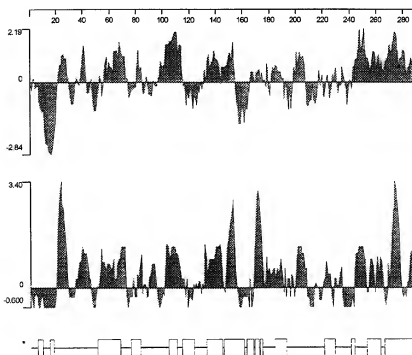
M1 ORF32



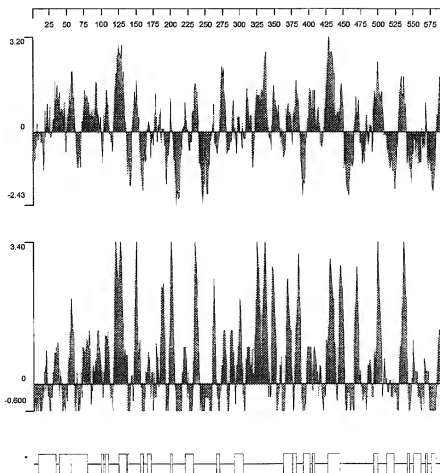
8/24

FIGURE 8

9/24

FIG. 8E**FIG. 8F**

10/24

FIGURE 9

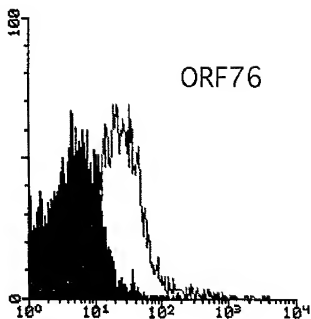
11/24

FIGURE 10**Fig. 10A**

M1 ORF76

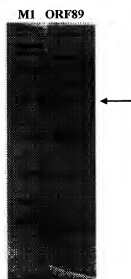
**Fig. 10B**

TP OMV

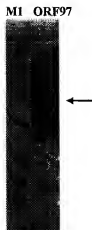
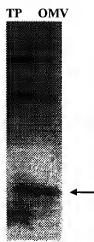
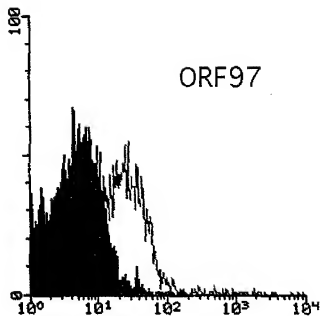
**Fig. 10C**

12/24

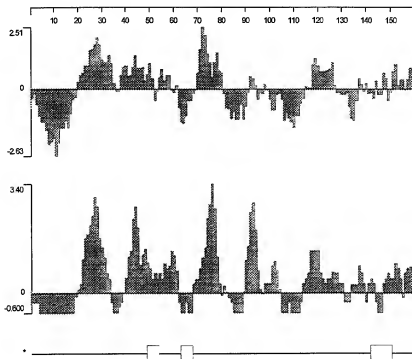
FIGURE 11



13/24

FIGURE 12**Fig. 12A****Fig. 12B****Fig. 12C****Fig. 12D**

14/24

Fig. 12E

15/24

FIGURE 13

Fig. 13A

M1 ORF106

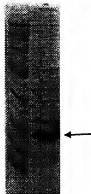
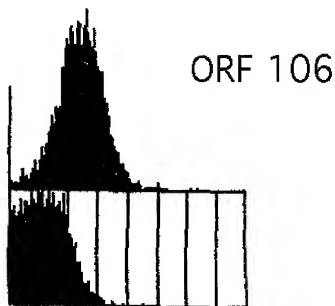


Fig. 13B

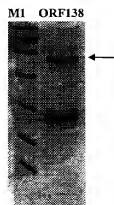
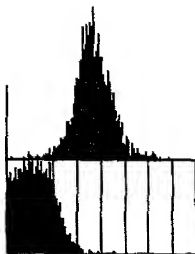
M2 ORF106



Fig. 13C



16/24

FIGURE 14***FIG. 14A******FIG. 14B***

17/24

FIGURE 15

Fig. 15A

M1 ORF23



Fig. 15B

M2 ORF23

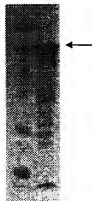
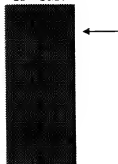
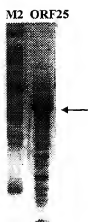
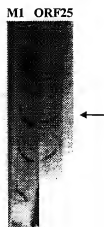
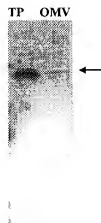
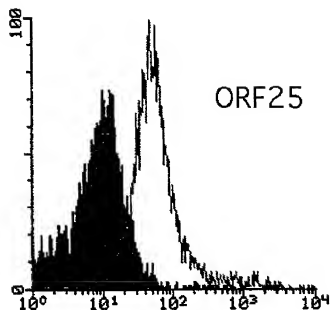


Fig 15C

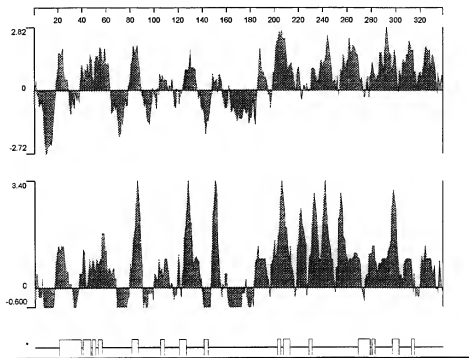
TP OMV



18/24

FIGURE 16**Fig. 16A****Fig. 16B****Fig. 16C****Fig. 16D**

19/24

Fig. 16E

20/24

FIGURE 17***Fig. 17A***

M1 ORF27

***Fig. 17B***

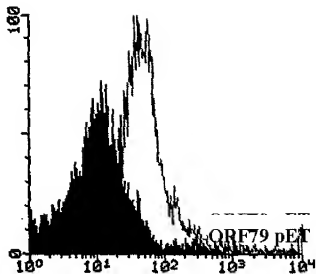
M2 ORF27



21/24

FIGURE 18***FIG. 18A***

M1 ORF79

***FIG. 18B***

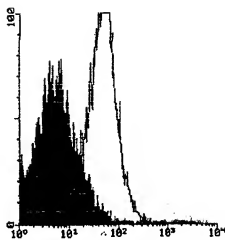
22/24

FIGURE 19**Fig. 19A**

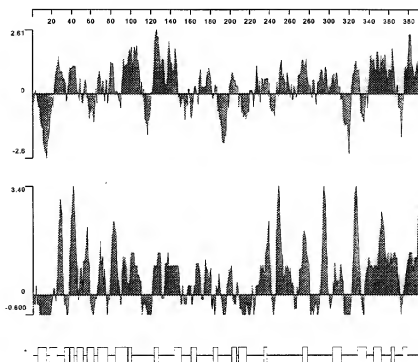
M1 ORF85

**Fig. 19B**

TP OMV

**Fig. 19C**

23/24

Fig 19D

24/24

FIGURE 20**Fig. 20A**

M1 ORF132

**Fig. 20B**

M2 ORF132

**Fig. 20C**

ORF132